A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data

Master's Thesis submitted

to

Prof. Dr. Stefan Lessmann

Humboldt-Universität zu Berlin
School of Business and Economics
Chair of Information Systems

by

Shikhar Srivastava
(570347)

in partial fulfillment of the requirements
for the degree of
Master's Program in Economics and Management Science (MEMS)
Berlin, February 24, 2017

Abstract

For photovoltaic (PV) based energy plants, early participation in energy auction markets, and efficient resource planning require accurate forecasts of solar energy. In this study, forecasting capability of a method of predictive learning called Long Short Term Memory (LSTM) has been shown in contrast to algorithms with a proven track record in the domain. The mesoscale models, thus compared, predict day-ahead Global Horizontal Irradiance, a proxy for energy produced through PV systems. These models are primarily built upon Global weather forecasts of atmospheric variables. Various locations are considered in order to conclude effectiveness of proposed algorithm. This comparative study uses remote-sensing data for testing the prediction accuracy at 21 locations, 16 of which are in mainland Europe and 5 in the US. As the compared algorithms learn better for a region with multiple locations than a single location, at all of these locations a small region of fake PV plants was created. A pragmatic approach of Transfer Learning was adapted to ensure minimal bias. Besides showing that the LSTM forecasts are better than others, the paper offers a novel method to the domain for such comparative studies in the future. Also, using both, LSTM and this design of experiment, together would function as a reliable tool for remote solar resource assessment.

Contents

List of Abbreviations

NWP	Numerical Weather Prediction	GHI	Global Horizontal Irradiation
AMS	American Meteorological Society	GFS	Global Forecast System
GEFS	Global Ensemble Forecast System	NN	Neural Network
GBR	Gradient Boosting Regression	ML	Machine Learning
CNN	Convolutional Neural Network	DL	Deep Learning
FFNN	Feed Forward Neural Network	RNN	Recurrent Neural Network
LSTM	Long Short Term Memory	BP	Back Propagation
RMSE	Root Mean Square Error	MAE	Mean Absolute Error
ReLu	Rectified Linear Unit	tanh	hyperbolic tangent

List of Figures

List of Tables

1 Introduction

Today, energy can be seen as the driving force of modern civilization as we know it. With accelerating growth of technological developments, we can only envisage this dependency on energy to be increasing. By 2035, the demand for primary energy and electricity is projected to increase by 50% and 70% respectively (Duffy et al. 2015). On the other hand, issues like environmental abuse due to traditional methods of energy production or struggles to meet demand against rapid population growth or international conflicts arising in context of fossil fuels create a distressing outlook of today's world energy. In response, the International Energy Agency (IEA) is aggressively pursuing research for renewable energy related technologies for a sustainable future. Growing awareness of the general public regarding afore-mentioned issues and acting as an environment friendly solution to these issues helped renewable energy resources gain vast popularity. However, to completely replace fossil fuel based energy systems, making these energy sources economically viable is a must.

Solar and wind energy are the two most popular form of renewable energy sources. Wind energy, if efficiently exploited, promises to meet the demands of energy. Whereas solar energy, being the most abundant source of energy on this planet, is heavily underutilized. The average solar power received by continents accounts to 23 million gigawatts (GW). This estimated value accounts for weather variation as well. For a relative comparison, a standard size nuclear power plant generates 1 GW. Solar energy abundance miniaturizes combined availability of all other energy sources (renewable or finite) such as wind, geothermal, hydropower, coal or nuclear (Perez et al. 2011). Countries around the world are actively seeking to capitalize on this free source of green energy. Between 2008 and 2015 the average cost of solar photovoltaic (PV) decreased by almost 80% (International Energy Agency 2016). In 2014, the IEA projected that by 2050 solar energy would constitute 27% of worldwide electricity consumption. Setting a motivating example, Germany, with considerably lower solar availability, has effectively integrated this energy into its electricity grid almost exclusively through installation and enhancements of PV based plants. In 2016, PV-generated power amounted to 38.3 Tera-Watt Hours (TWh) and covered approximately 7.4% of Germany's net electricity consumption. Germany plans to install 150-200 GW capacity through PVs by 2050 which translates to 4-5 GW PV installations every year (Wirth 2017). By 2020, most of the developed and developing countries are expected to have PVs contributing a significant source of energy (International Energy Agency 2016). Around then, PV is also expected to be competitive in five of the largest energy markets of the European Union (EU) attributing to the factors such as increased PV efficiency, economies of scale, and the development of markets (Hoffmann 2011). The biggest hassle of shifting to this energy source is its high variabil-

ity and uncertainty in both time and space. Accurate forecasting of energy availability is, therefore, industry's top priority.

In most of the countries, energy markets are divided in day-ahead and intra-day market sessions. In European Power Exchange Market, for example, power trading is organized in various time horizons of which day ahead and intra-day are mostly used for energy auctions. Similar horizons are used in the US energy markets as well. For solar plants connected to the grid, it is necessary to supply a required amount of energy. Based on this requirement, relevant authorities of power plants need to make decisions regarding energy purchases based on whether the required quota is predicted to be met or not. Expectedly, the sooner a commitment is made the lower the prices are, and larger the forecast error is, larger will be the cost of energy as the market follows a bid system. Due to unavailability of PV data, scientists tend to perform PV energy predictions via forecasting availability of solar irradiance flux instead.

On the other hand, the recent advancements in predictive modeling techniques using machine learning (ML) based algorithms are enabling us to better predict complex natural phenomenons. It is of obvious interest to researchers and scientists to be able to find hidden non linear relationships between observations to understand their behavior. Even though there exists a vast literature on irradiance forecasting, it can be observed that there is a lag (in time) in adaptation of new/updated techniques from the ML community. This paper explores a relatively new algorithm of prediction, Long Short Term Memory (LSTM), which belongs to the family of prediction techniques called Deep Learning (DL), for solar energy forecasting. The experiment performed comprehensively checks the efficiency of the method for a day-ahead forecast horizon.

This paper discusses the application of LSTM in irradiance forecasting and provides a base for future developments and applications in other domains of meteorology. A short summary of all methods of irradiance forcasting can be found in section 2. Followed by a review of comparable literature work in section 3. The basic methodology and related intricacies of LSTM is provided in section 4 and a brief on compared algorithms is provided in section 5. Section 6 explains the experimental setup and underlying motivations. Results and discussion can be found in section sections 7 and 8. Conclusion of this experiment along with proposed future work is provided in sections 9 and 10.

2 Irradiance Forecasting: An overview

Global Horizontal Irradiance (GHI) is the total flux of downward short-wave radiation on a horizontal surface. For photo-voltaic cells, GHI value is of primary concern as the energy produced from them can be directly calculated from it. Extraterrestrial irradiation (also known as clear-sky irradiation) is radiant flux on the atmosphere (before interacting with it). It can be calculated for any place using earth's position relative to sun. The amount of GHI arriving on planet's surface is largely filtered due atmospheric conditions. As it passes through the air molecules, water vapor, aerosols, and clouds, it gets either scattered or absorbed. The high variability of these elements and our incomplete understanding of their formation/depletion behaviour in time and space raises the issue of uncertainty in solar energy generation.

As mentioned by Kostylev and Pavlovski (2011), irradiance forecasting horizons can be broadly divided into four prediction problems depending upon industry requirements. (1) A very short horizon based forecasting or "now-casting" allows operators in optimized market clearing and regulating source alteration. Nowcasting's time horizon ranges from 1 to 30 minutes. (2) For economic intra-day operational planning which includes programming backup, load increment/decrement, urgent purchases and peak load matching, predictions ranging from 30 minutes to a few hours assists. This is often referred to as short-term forecasting. (3) Mid-range forecasting, a day to a week ahead, are essential for power producers in decisions regarding early market participation (unit commitment) and scheduling of maintenance. (4) Finally, long-range predictions, which range from a week ahead to an year, are mainly useful for solar resource assessment. In the literature, forecasting models have been divided into sub-categories on the basis of their nature. An elemental understanding of all of them is required for comprehension of on-going research in the domain.

2.1 Stochastic Models

Traditionally, statistical regression methods were used for univariate time series based irradiation prediction where the input data would be taken from dedicated ground based radiation recording instruments. In literature, various versions of Autoregressive Integrated Moving Average (ARIMA) have been researched and found useful. First applied by Chowdhury (1990), the linear stationary models were the first ones to show high sub-hourly forecast accuracies. Inman et al. (2013) provided a comprehensive look into many techniques and a comparison of many of these models can be found in a study by Reikard (2009) which showed that multi-variate input based regression (including precipitation, cloud cover, wind components etc.) produced more accurate predictions.

Elizondo et al. (1994) and Mohandes et al. (1998), were the first ones to propose application of non-linear learning algorithms for irradiance prediction problem. In principal, these algorithms strive to find relationships in data hidden from human eye, and the most widely used and researched algorithm in the domain is Feed Forward Neural Networks (FFNN). Researchers have shown different variations of this method being capable of predicting irradiances with horizon of minutes to a few hours ahead. The growth of computational power has made the use of their application and the manipulation of their structure easier. As a result, there has been a significant increase in application of various versions of such learning algorithms (Mellit and Kalogirou 2008; Mellit et al. 2009). However, exclusive ground based data modeling saw a decline due to the growth of other more sophisticated methods of weather monitoring technology which decrease surface based instrument installations. On the other hand, these models are being heavily used as an integral process of other irradiation modeling techniques for obtaining a higher accuracy. I have discussed this in detail in section 2.4.

2.2 Satellite Models

The major source of uncertainty in irradiation forecasting can be attributed to not knowing the future cloud cover and its depth at a given time and location. Instant snapshots of atmosphere are being taken in real-time by geo-stationary weather satellites covering vast spatial resolution. Irradiance can be directly calculated using the Heliosat method, which was first proposed by Cano et al. (1986) and further improved by Beyer et al. (1996); Hammer et al. (2003). The underlying method includes calculation of cloud density using the radiation amount reflected back from earth. This reflected radiation is the sum of reflected radiation from the ground and the clouds. It is decomposed and further processed in accordance with Hammer et al. (2003) to obtain irradiance on earth. Another, and more accurate, way of doing it is by calculating cloud motion vector fields in accordance with method proposed by Beyer et al. (1994) and later revised by Lorenz et al. (2004), which accurately predicts irradiance incident on earth's surface. The successive cloud-index images, which are taken from ground and satellite over a few hours explain their spatial coverage and velocity. Time series models are also applied for detecting motion vector fields to extrapolate cloud behavior in higher temporal resolution (Hammer et al. 1999). As a part of post-processing, a smoothing filter is applied on each pixel extrapolated image by averaging $a \times a$ pixels in the area (Kühnert et al. 2013). This post-processing reduces inaccuracies and allows prediction of irradiance for up to 6 hours with high accuracy[1]. Weather forecasts use results of these models for a horizon of 5-6

[1] For more details on derivation, specificities and thorough review of Satellite based models one can refer to Lorenz et al. (2004); Lorenz and Heinemann (2012)

hours till the temporal change in cloud structure can be derived from its motion, after which their efficiency declines.

Often criticized, the method of simple temporal and spatial extrapolation is the major contributor for mid and long-range forecast errors. However, it is unavoidable as dynamic intricacies of weather behavior cannot be captured by atmospheric snapshots. A better understanding of variables' interdependencies and their physical properties is essential to account for their intermittent nature. Variables in this context means various physical processes. Irradiance for all sky conditions including cloudy skies may be derived using radiative transfer models (RTM) (Heinemann et al. 2006), requiring input on the vertical structure of cloud physical parameters, for e.g. cloud and ice water content or droplet radius. This is called meso-scale modelling which comes under a broader segment of Numerical Weather Predictions.

2.3 Numerical Weather Models

For forecasts beyond a horizon of six hours Numerical Weather Prediction (NWP) models are used in practice. NWP models marked their origin in early 1920s and over the last half century, NWP domain witnessed immense growth owing to evolution of super-computational abilities[2]. What makes NWP different from statistical models is that they are essentially physics based models which require an amalgamation of various differential equations characterizing dynamics of atmosphere. The atmospheric equations are derived from different field of physics such as fluid/gas dynamics, heat transfer theory etc.[3]

2.3.1 Global Forecasts

As an initial input, observations are taken from satellites and ground stations. This data represents atmosphere and is used in the physics based differential temporal equations to model the dynamic environment. Next, the model outputs are resolved on "mesh-grid" using numerical methods and spatial discretization defining the forecast region. The forecasts of these models are not only for surface irradiation but also other atmospheric variables like humidity, temperature and wind components. The resolution of this forecast region depends on the scale of resolution of satellite images. Global forecasts usually have coarse resolutions and are not suitable for getting specific point location's atmospheric details. However, these resolutions have seen significant improvement over the last few years. For example, Europe's ECMWF is able to provide a resolution of

[2] See Lynch (2008) for a comprehensive view on NWP's history

[3] For an in-depth understanding of these equations and methods one can refer to Eugenia Kalnay (2003)

$0.125° \times 0.125°$(~$16 \times 16km$) since 2010 and USA's Global Forecasts System (GFS) is able to provide $0.5° \times 0.5°$ since 2010 (which was $1° \times 1°$(~$100 \times 100km$) since the early 1990's). There exists 15 such systems worldwide from different participating countries. Usually, the outputs of these models are in timesteps of 3 hours and extends up to several days (Lorenz and Heinemann 2012).

2.3.2 Mesoscale Forecasts

Accurate point specific irradiation prediction is the final aim of the whole domain. Various methods have been developed by weather research departments of different countries for small scale feature mapping from Global Forecast's data. In general, all these models use initial and boundary conditions from Global NWPs (Lorenz and Heinemann 2012). The difference usually lies in the input data which parameters are given importance, and in the incorporation of local terrain for spatial averaging. The outputs of these models are hourly which requires a temporal interpolation as Global forecasts are for fixed timesteps. Accuracies of these models are better for mid-range horizons. A few examples of such models are GFS based Weather Research and Forecasting (WRF) model (Skamarock et al. 2005) or fifth generation mesoscale model (MM5) created at Pennsylvania State University and the National Center for Atmospheric Research (Dudhia et al. 2005) or ECMWF-OL, a mesoscale model created using ECMWF data by researchers of University of Oldenburg, Germany (Lorenz et al. 2009a,b). There are many others, covering explicit details of these models will be too broad for the scope of this report. For review on most of such methods see Espinar et al. (2010) and Diagne et al. (2013), and benchmarking details can be found in a comparative study done by Perez et al. (2013, 2010) and Müller and Remund (2010).

It is important to note that the chief source of errors in mesoscale NWPs propagates from poor resolution associated with parent Global Forecast model. Almost everywhere in the literature ECMWF's related models (like ECMWF-OL) outperform other models, which, apart from the methodology involved, should also be attributed to better resolved ground level readings of ECMWF's global forecasts (if it is used). On the other hand, the relative unforced errors (which are not due to parent Global Forecast values), comes from oversimplified generalization of terrains of area in focus for which mesoscale model is built.

2.4 Hybrid Models

Section 2.1 summarized irradiation prediction methods using ground based readings as inputs in linear and non-linear statistical algorithms. Hybrid Models are similar stochas-

tic techniques but built on a broader range of inputs and use computational systems to perform predictive learning. Ground readings taken from instruments are often referred to as endogenous inputs in literature. The Hybrid models are so called as they can be applied over outputs of Global or Satellite models, which are called as exogenous inputs. When coupled with on-going evolution of computational abilities and increased accuracy of GFS forecasts, this flexibility has created a surge of research over the last decade. Similar to ground-based applications, when non-linear models are considered neural networks grabbed the attention of most of the researchers. As can be inferred from the summary tables of reviews specific for such models (Diagne et al. 2013; Inman et al. 2013; Raza et al. 2016), most of the predictions aimed for a low horizon period (from minutes to a few hours). It could be attributed to non-availability of ground-based irradiation readings for longer durations. Those readings, in principal, are required for checking a model's performance.

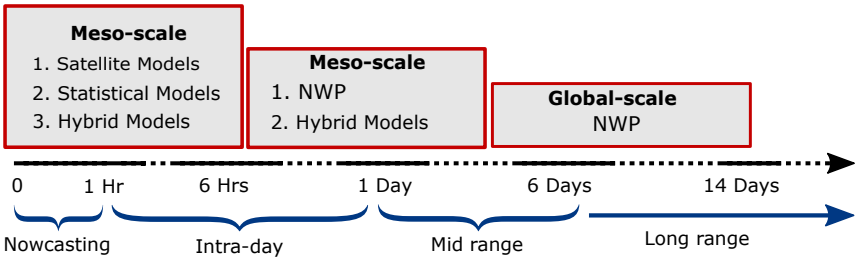


Figure 1: Models used for Irradiance forecasting at various horizons

The Hybrid models in this context is used for applied machine learning (ML), a field of predictive modeling heavily integrated with computational sciences. To summarize, figure 1 gives a good picture of industry usage of all the models for different forecasting horizons. Hybrid models have shown to be efficient in forecasting a horizon of very short, short and mid ranges. Even some of the NWP mesoscale models use ML methods to improve their prediction accuracies. A few examples of such models in-use would be BLUE Forecast, a mesoscale NWP model built upon GFS outputs by Austrian Meteorological department, is integrated with statistical post-processing procedure using different methods of data mining like ridge regression, automatic quadratic models or neural networks (Perez et al. 2013) or ECMWF-Ol which also uses neural networks for post-processing NWP outputs (Lorenz et al. 2009b).

For this experiment, efficiency of state of the art ML algorithms have been tested for day-ahead irradiance forecasting. The models were built using the input data from Global Forecasts and target data from Satellite models. The corresponding literature has been discussed in the next section.

3 Related Work

Irradiance forecasting does not follow a one-size-fits-all approach. Section 2 explains that for various forecast horizons different models are preferred. For an objective and comparative literature study, I have selected only works which had a forecast horizon of at least 1 day. Table 1 is a comprehensive list of most of such major published studies predicting irradiance using ML algorithms. A thorough review of all other forecasting methods has been provided by Diagne et al. (2013), Inman et al. (2013) and Raza et al. (2016).

Given the nature of the irradiation forecasting problem along with the degrees of freedom of designing an experiment, most of the forecasting studies done are incomparable. Underlying reason for this incongruity is also evident in Table 1, every study performed was based on specific spatio-temporal related contextual data. The unique characteristics of weather in both space and time makes replication of results impossible for a different location, i.e. the models built were applicable only for the location of study. Hence, direct comparison of error values from the results of other studies should be considered redundant unless they are for the exact same time and place. Nevertheless, these studies are informative for their innovative designs and end-results. For example, Cao and Lin (2008) proposed Diagonal Recurrent Wavelet Neural Networks for effective capturing of non-linearity in time-series structure of weather related variables for forecasting irradiance. Linares-Rodríguez et al. (2011), after getting access to radiation data of 83 stations of a small coastal region in Spain, built a regional model by training neural networks on 65 of them and testing on the remaining 18. Whereas, Moreno et al. (2011) built a neural network based model for the whole country of Spain using ground readings recorded at 40 stations that were sparsely distributed in the country.

The success of ML algorithms in various regression/classification tasks has been a clear result of boosted computational powers and data storage systems. Coupled with increasing access to weather data (like free-sourced NOAA's GFS), a vast pool of opportunities has been unveiled for ML applications and, hence, the number of related studies in the last few years has increased (Voyant et al. 2017). Table 1 shows how application of Feed Forward Neural Networks(FFNN) dominated the irradiation forecast domain. In contrast, a supervised ML algorithm called Gradient Boosting Regression (GBR), which has received little attention in the domain, produced the best predictions in a solar-energy prediction competition hosted on Kaggle[4] in 2014 (McGovern et al. 2015).The competition was hosted by American Meteorological society (AMS) and provided a huge dataset containing 14 years of temporal records and PV output values from sporadically spread

[4] Kaggle is an online platform for data mining competitions. Competition website: www.kaggle.com/c/ams-2014-solar-energy-prediction-contest

98 PV plants covering the whole region of Oklahoma. The Global Model's forecasted atmospheric variables at different time steps of a day were the provided inputs. Solar energy data of such large spread in both time and space has rarely been researched before. A study performed by Lorenz et al. (2009a) is one of the very few cases. The study constituted mesoscale NWP modeling using 200 meteorological stations data from all over Germany. A novel approach of pixel based spatial averaging was derived from that study. Likewise, the outcomes of Kaggle's competition were surprising for the atmospheric science community. The competition results unanimously proved that a higher accuracy can be achieved via GBR as almost all of the top scorers used this for final predictions. GBR won over traditional algorithms like FFNN or Linear Regression which were also applied by some participants (McGovern et al. 2015). This outcome revealed a major gap in mid-range irradiation forecasting which pushed researchers to explore supervised learning algorithms. For example, Support Vector Regression (SVR) method for irradiance forecasting was found to be efficient by Fonseca J. et al. (2014) and Ekici (2014). In large number of research of applied machine learning methods, ensemble models have a track record of surpassing accuracy. Ensemble models were suggested in this domain by Zhaoxuan et al. (2016) and Gala et al. (2016). The former used hybrid of FFNN and GBR, whereas, the latter showed that superior models could be made over SVR, GBR and Random Forest Regression (RFR) by combining all of them.

A recent review by Voyant et al. (2017) has been published exclusively for irradiance forecasting using machine learning methods. At first glance, it is surprising to find only a handful of published research papers, which have focused on a day-ahead or longer forecast horizons. However, taking a closer look on the evolution of experimented algorithms, a speculative reasoning appears to be that the initial success of non-linear algorithms in forecasting short-term horizons and failure in long range motivated most ML researchers to concentrate on improving accuracy in short horizon segment. This was either done via tweaking modeling structures or by adding more endogenous/exogenous inputs. Improving on such well established grounds would allow one to contribute in accuracy improvements and provide a direct comparison with other state of the art methodologies. Another reason for such manifestation and a more prominent one is that ground level irradiance readings which are used to test model performance were available (or manually obtained) only for sites in focus. Besides ground based irradiance readings, researchers used direct PV output provided by local solar energy companies which is not always obtainable due to data policies/logging practices. Despite repeated conclusive statements made by NWP researchers that highly accurate satellite irradiance data can be used as a proxy ground observed irradiance (Lorenz et al. 2009b; Lorenz and Heinemann 2012; Tadesse et al. 2016; Sengupta et al. 2015), none of the ML researcher from table 1 capitalized this information in their experiments.

Multiple insights can be drawn out from table 1. First and foremost, Deep Learning (DL) based methods like Convolutional Neural Networks (CNN) or Long Short Term Memory (LSTM) have not yet been tested in the irradiation forecasting domain. This has also been confirmed in the conclusion section of Voyant et al. (2017). Also, little to no research has been done using other popular machine learning algorithms such as xgboost or Random Forest or Bayesian learning methods. Second, as mentioned above, none of the researchers have performed their tests in more than one country because of their dependency on ground readings of irradiance. This also resulted in almost no inter-climatic validations of models thus produced. The models thus created would only explain the irradiation behavior of research of a specific place on earth and not any other. Third, for many studies, testing of models was performed on selective months of the year or just a few days. This creates an unavoidable problem of selection bias. Validation on all seasons is desirable for generalizing the efficiency of models. Finally, all the studies with input data type of hourly or higher resolution concentrated on prediction accuracies of intra-day horizons. A day ahead prediction was provided only as an extended result.

Besides, there are always valid dangers of over-tuning of proposed models as mentioned by Hand (2006). Often, in comparative studies, researchers tend to focus on outputs of the algorithms they are proposing. Coupled with the above-mentioned problem of location specific studies, we can expect many results to be biased towards proposed methodologies. In this experiment, I have tried to take into account almost all of these research gaps and model generalizing issues via a pragmatic design of experiment. The issue of location and climate specific modeling has been solved by eliminating ground measurement based dependencies and by doing so, being able to perform experiment at multiple locations on earth. An entire year has been taken as the validation set which checks models' applicability over all seasons. Furthermore, this is an exclusive day-ahead prediction study with motivation for minimizing loses via efficient participation in day-ahead auctioning in energy markets. Also, a simple procedure of Transfer Learning has been used for validating models at multiple locations, which, in effect, reduced the looming dangers of over-tuning any of them. Finally, since it is not feasible to test all the top performing predictive learning algorithms, for this experiment only long short term memory (LSTM) has been tested and compared against some other machine learning algorithms which have a proven track record.

	Geographical Span			Data			Predictors			Method***	
Study (in chronological order)	**Countries covered**	**Distinct locations**	**Climates covered***	**Training set size****	**Test set size****	**Type**	**Geospatial/ Temporal**	**Ground recorded**	**Weather forecast**	**Proposed**	**Compared**
Kemmoku et al., 1999	Japan	1	1	1827	365	Daily		x		FFNN	Persistence
Podestá et al., 2004	Argentina	1	1	365	1276	Daily		x		GAM	None
Yona et al., 2007	Japan	1	1	333	4	Hourly		x	x	RNN, RBFNN	FFNN
Cao and Lin, 2008	China	1	1	7992	744	Hourly		x	x	DRWNN	AP, FFNN
Bacher et al., 2009	Denmark	1	1	1460	NA	Per 15 min			x	RLS	Persistence
Paoli et al., 2010	France	1	1	6574	547	Hourly		x		FFNN	NP, MC, K-NN, AR
Linares-Rodríguez et al., 2011	Spain	83	1	219439	60768	Daily	x		x	FFNN	None
Marquez and Coimbra, 2011	USA	1	1	236	158	Daily	x		x	FFNN	NWP, Persistence
Ding et al., 2011	USA	1	1	7	2	Per 30 min		x	x	FFNN	None
Chen et al., 2011	China	1	1	670	74	Hourly			x	RBFN	None
Moreno et al., 2011	Spain	40	3	1241	219	Daily		x		FFNN	BC, KRR
Cococcioni et al., 2011	Italy	1	1	359	6	Per 15 min		x		FFNN	None
Ekici, 2014	Turkey	1	1	1096	365	Daily		x		SVM	None
Fonseca J. et al., 2014	Japan	4	2	1464	1460	Hourly	x		x	SVM	Persistence
Voyant et al., 2014	France	1	1	2922	731	Daily		x		FFNN	ARMA, Persistence
Salcedo-Sanz et al., 2014	Spain	1	1	292	73	Daily	x	x	x	Hybrid ELM	ELM, SVR
Long et al., 2014	China	1	1	723	38	Daily		x		None	FFNN,KNN,MLR
Fernández et al., 2014	Spain	1	1	365	181	Daily			x	SVR	None
Cornaro et al., 2015	Rome	1	1	876	584	Per min		x	x	FFNN	Persistence
Zhaoxuan et al., 2016	USA	1	1	365	NA	Daily		x		FFNN	SVR
Gala et al., 2016	Spain	7	2	365	302	Daily			x	Ensemble	SVR, GBR, RFR
Wang et al., 2016	China	1	1	365	181	Daily		x		PF-LR	FFNN, RBF
Leva et al., 2017	Italy	1	1	120	30	Daily		x	x	FFNN	None

* Variation in climate follows Koeppen Climate Classification System (Peel et al. 2007)

** Training and test sizes are number of distinct days of taking the observations regardless of further granular measurements

*** Table 9 in appendix expands the model abbreviations used here.

Table 1: A comparative list of studies done for forecast horizon of at least 24 hours

4 Recurrent Neural Networks

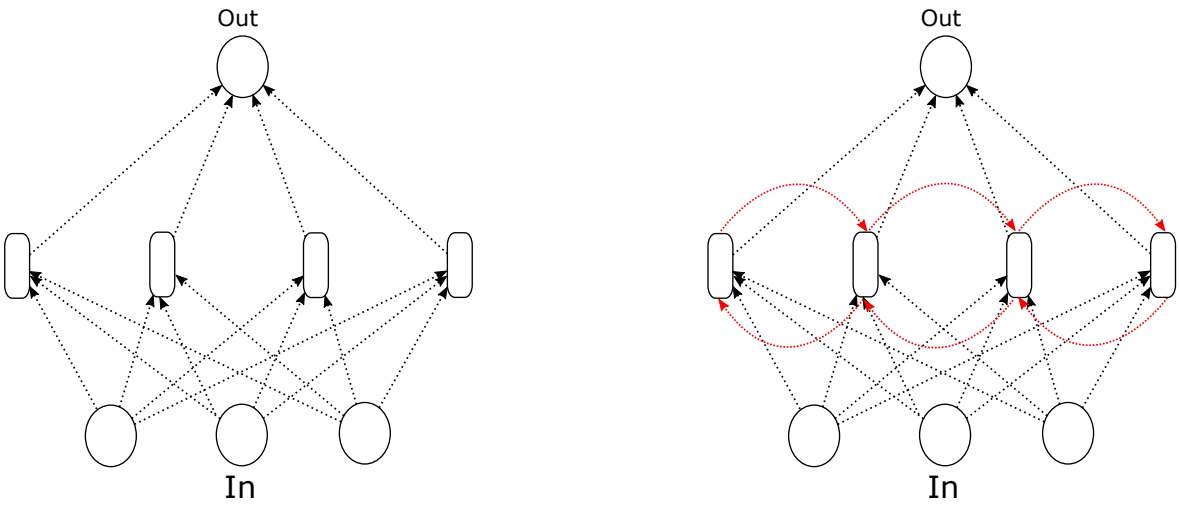
This paper describes experimentation and application of Long Short Term Memory (LSTM) algorithm in forecasting solar irradiation. The algorithm belongs to a family of neural network(NN) architecture called Recurrent Neural Networks (RNNs). Elemental understanding of both will be helpful for future treatment and development in similar tasks. A short summary on groundwork of RNNs has been discussed in section 4.1. The basic underlying architecture and problems of parent architecture is given in 4.2, and a detailed explanation and working of proposed algorithm is provided in 4.3.

4.1 Background

An effort towards creating artificial intelligence, NNs were conceptually developed in the early-1940s inspired by the interactive design of nervous system in human brain. The major improvements in working back-propagation (BP) algorithms for neural nets were not done till the 1980s (Werbos 1981; Rumelhart et al. 1986; Le Cun et al. 1988). Soon after which Hecht-Nielsen (1989) and Hornik et al. (1989) proved that a NN system with a single hidden layer of neural cells can approximate any continuous function, which means that the system can learn highly non-linear relationships in data. Based on this theorem researchers extensively used NNs in pattern recognition ever since. A subdivision of NN architecture called Recurrent Neural Networks (RNN) using stochastic gradient descent for BP was also created in the 1980s (Rumelhart et al. 1986). The underlying motivation behind RNN was to enable self-organization of synapses in NNs. The parent, and now the other subdivision, being feed-forward neural networks (FFNN) worked on a simple mechanism of pushing the inputs through a network of layers with neurons containing an approximating function and revising their values using BP. Inherent difference between FFNNs and RNNs is that the former do not have a feedback (or iterative) loop. figs. 2a and 2b shows a schematic diagram of each. One can refer to Schmidhuber 2015 for a comprehensive summary of the ongoing evolution of neural network systems.

4.2 Methodology

In addition to functional properties of a typical FFNN, the fundamental characteristic of RNN is that it has iterative loops of the neuron units in hidden layers (figure 2b). This architecture allows the approximated values of hidden units to share parameters or, in other words, learn from each other. The sharing of parameters can also be considered as building a "memory", allowing the system to, in theory, remember and hence, predict sequences. As compared to FFNN, RNN has a closer linkage to the functioning of



(a) A simple FFNN (b) A simple RNN

Figure 2: Bifurcation of Neural Networks

synapses in a brain. A human-brain learns most of the things in progressive order rather than randomized for eg. phone numbers, alphabets, songs etc. Additionally, it has been shown that RNNs are Turing-Complete (Siegelmann and Sontag 1992) i.e. they are computationally universal. The sequential nature of learning can be better understood via its simplification to a 1-D temporal array of inputs $x_1, x_2, ..., x_t$ where x is input at different time slices $1, 2, ..., t$. The parameter sharing across distinct time indices allows it to generalize and share statistical properties over various positions in the sequence. Figure 3 represents a simple RNN architecture[5] with 1 hidden layer h and output y, implying x maps to h, h maps to y and h also updates with its past-self.

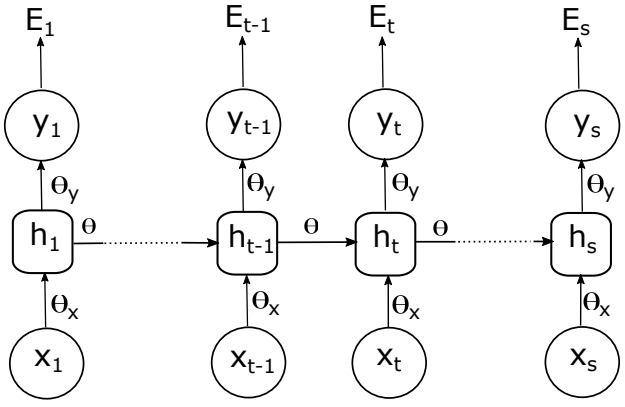


Figure 3: A single layered RNN

This self-recursion can continue till infinitum. The dynamic system at state t can be represented by equations 1 and 2.

[5] This is a simplified version of Forward Pass RNN. There exists many variations, a comprehensive overview of which is provided by Goodfellow et al. 2016.

$$h_t = \theta\phi(h_{t-1}) + \theta_x x_t \tag{1}$$

$$y_t = \theta_y \phi(h_t) \tag{2}$$

Parameter θ_x and θ_y represents weights of inputs and outputs at time t. Weight of the hidden units, parameter θ, is shared across the hidden unit and, hence, stays the same across the network (which also saves computational memory). ϕ is user-specified activation function applied for approximations. The losses at each time step, as shown in figure 3, is calculated as E which is backpropagated at each t over horizon S for finding the global minima via equation 3.

$$\frac{\partial E}{\partial \theta} = \sum_{t=1}^{S} \frac{\partial E_t}{\partial \theta} \tag{3}$$

$$\frac{\partial E}{\partial \theta} = \sum_{k=1}^{t} \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial \theta} \tag{4}$$

Because the parameter θ could have occured before time t, it is required to sum all the instances k before time t. Equation 4 represents the back-propagation using the chain rule. This, in theory, exploits the temporal or spatial dependence of data, and proves to be a major asset of RNNs in utilizing contextual information. However, in practice, researchers could not apply it for learning from deep sequences or from large contexts because of the issue which was not identified until early 90s.

4.2.1 Vanishing/Exploding Gradient Problem

Hochreiter (1991) zeroed down the fundamental cause of inefficiency in RNNs. From (4), it can be promptly inferred that vanishing (or exploding) of local error would lead to vanishing (or exploding) of global error. This complication is produced from jacobian $\partial h_t / \partial h_k$ in the chain of (4), which is a product of jacobians (b in (5)) going layer by layer from t to k. Using (1), this product gives d in (5) where ϕ' is the derivative of ϕ. To account for all dimensions diagonal matrix form is used (c in (5)).

$$\underbrace{\frac{\partial h_t}{\partial h_k}}_{a} = \prod_{i=k+1}^{t} \underbrace{\frac{\partial h_i}{\partial h_{i-1}}}_{b} = \underbrace{\prod_{i=k+1}^{t} \theta^T \underbrace{diag[\phi'(h_{i-1})]}_{c}}_{d} \tag{5}$$

As have been thoroughly explained in Bengio et al. (1994) and Hochreiter et al. (2001), if the value of leading eigen-value in weight matrix or $\|c\| > 1$ then the product increases exponentially forcing $\|a\| \to \infty$ and hence, blows up the error. Also, if $\|c\| < 1$ then $\|a\| \to 0$ which vanishes the gradient. In both complications it becomes impossible for RNNs to learn from $k << t$ or deep contexts.

4.3 Long Short Term Memory (LSTM)

To specifically address vanishing/exploding gradient problem (4.2.1) Long Short Term Memory, a cleverly modified RNN, was proposed by Hochreiter and Schmidhuber (1997) and further improved by Gers et al. (2000, 2002). LSTMs are recurrent systems but with self-connected "gates" in the hidden units. When working over deep contexts or long periods of time, these gates allows LSTM units to read, write and forget. This characteristic enables the units to hold only relevant data and erase the rest, and hence, sustain a constant error. Retaining only selective information brings the LSTM architecture much closure to human brain analogy. For eg. selective memories are stored in the brain rather than all events of one's past. LSTM and its adaptations have shown potential of solving a range of problems involving deep context based sequential learning. As have been compiled in Graves (2012, chap. 4), LSTM has proved to be highly efficient in learning context free languages (Gers and Schmidhuber 2001), recalling high precision real numbers over extended noisy sequences (Hochreiter and Schmidhuber 1997) and various tasks requiring precise timing and counting (Gers et al. 2002). Gaining advantage over traditional RNNs with its ability of remembering long sequences and to maintain stable error has made it state of the art in solving many AI related tasks. LSTM has accomplished success in many real-world situations such as protein secondary structure prediction (Chen and Chaudhari 2005; Hochreiter et al. 2007), music generation (Eck and Schmidhuber 2002), speech recognition (Graves et al. 2005) and handwriting recognition (Liwicki et al. 2007) among many others. In recent years, there has been a dramatic increase in LSTMs development and usage. The increase can be affirmatively attributed to evolution of GPU based data processing units allowing efficient computation of extensive number of hidden units embedded in multi layered networks.

4.3.1 Working

To understand the functioning of LSTM memory units, we can look at a single localized LSTM cell in the first layer of a network at a timestep t. Fig.4 represents this in which input (i_t), update (g_t), output (o_t) and forget (f_t) represents gates' output values. All the gates get an input of previous LSTM unit's output (h_{t-1}) and input data of current timestep (x_t). This is represented in the figure using color coded arrows. (6)−(11) are governing equations which are explained below. Please note that in these equations θ_{xn} and θ_{hn} are weight matrices of input data and output from previous cell whereas b_n represents bias vectors with $n \in (i, g, f, u)$, which are aforementioned gate nomenclatures.

Forget gate, which is the key element of LSTM architecture, chooses how much infor-

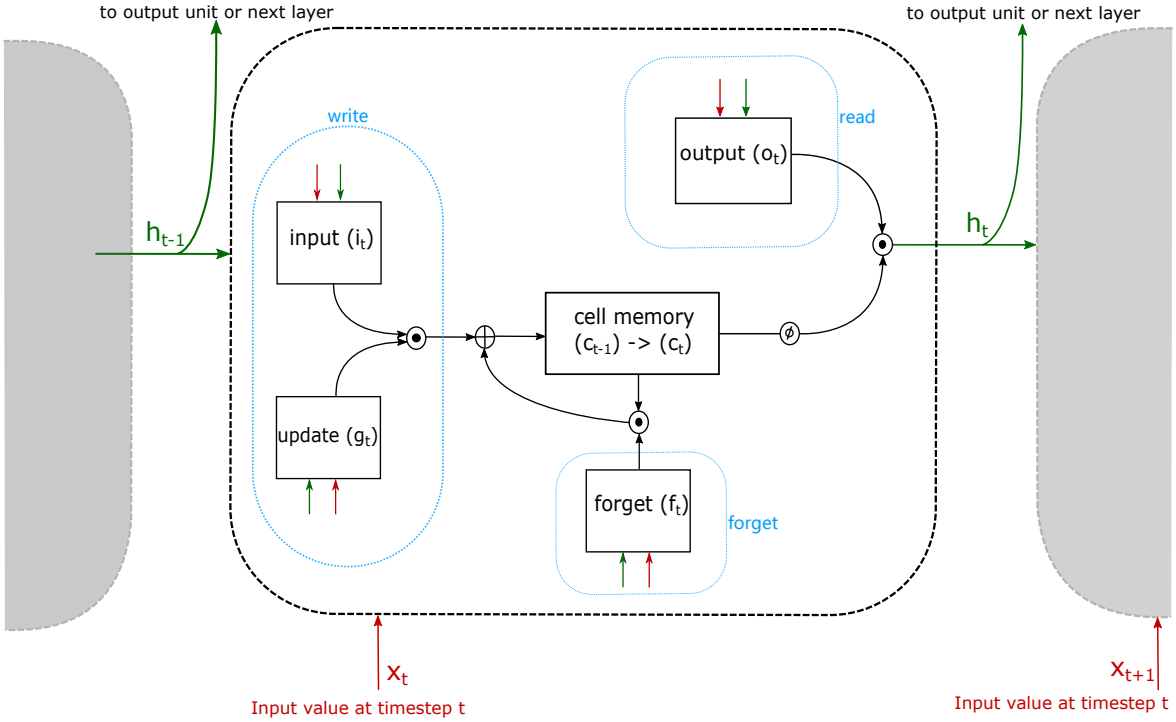


Figure 4: An LSTM unit in 1st layer. Symbol $\odot$ represents point-wise scalar multiplication and $\oplus$ is sum function. Color coded arrows shows direction of inputs to a system. The cell output activation function, represented by $\varnothing$, inside a O is user-defined (which usually is *tanh* but in our experiment it is *ReLu*.

mation should be deleted from memory. It operates through a *hard sigmoid* activation (linearly approximated standard *sigmoid*) which has a range of 0(forget all) to 1(remember all) and is governed by (6).

$$f_t = sigm(\theta_{xf}x_t + \theta_{hf}h_{t-1} + b_f) \tag{6}$$

The interaction (element-wise multiplication) of input and update gate (refer figure 4) performs write function in the memory. The input gate, using *hard sigmoid*, decided which values it wants to write and the update gate, using *hyperbolic tangent* (*tanh*) activation, creates a vector of new cell values. (7) and (8) represents their working.

$$i_t = sigm(\theta_{xi}x_t + \theta_{hi}h_{t-1} + b_i) \tag{7}$$

$$g_t = tanh(\theta_{xg}x_t + \theta_{hg}h_{t-1} + b_g) \tag{8}$$

Cell memory (refer figure 4) updates itself recursively by interaction of its old value $(t-1)$ with forget and write gates' values. It follows (9)

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{9}$$

Finally, the output gate, which serves a function of read, is combined (element-wise multiplication) with cell memory to calculate output of the cell h_t, also referred as cell state. An activation function, ϕ, is performed on cell memory for bounding the output values, and determines which values would go in the output from the output gate, o_t. Activation function, ϕ, in cell output equation is usually $tanh$ but for this experiment *Rectified Linear Unit* ($ReLu$) was the optimal choice. Equation (10) governs the read function of output gate o_t and (11) calculates the final cell output.

$$o_t = sigm(\theta_{xo}x_t + \theta_{ho}h_{t-1} + b_o) \tag{10}$$

$$h_t = o_t \odot \phi(c_t) \tag{11}$$

4.3.2 Regularization

LSTMs or deep neural networks, in general, are highly prone to over-fitting. As the network of LSTMs learn, weights adjust so much as to only be able to explain behavior specific of training data. A recently popularized method of tackling over-fitting, "Dropout" regularization, which is explicitly built for deep NNs (Srivastava et al. (2014)) was used in this experiment. Dropout process blocks some random set of cell units over each iteration. The number of of cells to be blocked (dropout layer value) is user-defined. It can be seen as an ensemble of various networks with each ensemble member being created in every iteration via bagging of the unblocked units. Figure 5, shows an instance (one iteration scenario), with a dropout layers applied after two fully connected NN layers.

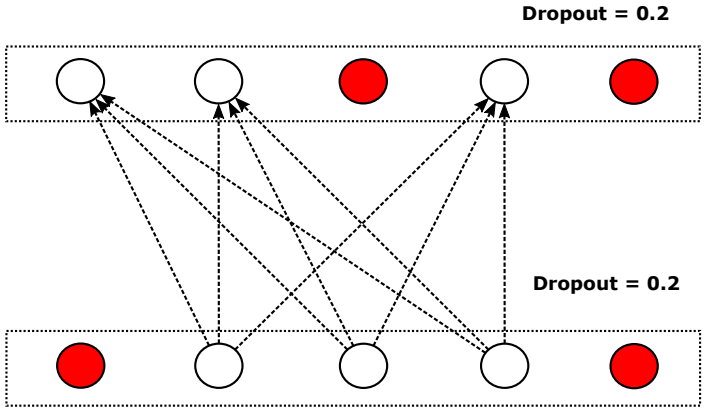


Figure 5: A dropout layer applied after two fully connected hidden layers

Final model averages the whole network with scaled down learned weights. This is similar to averaging technique used for creating ensemble models.

4.3.3 Weight Initialization and Optimizer

Both weights initialization and selection of optimizer are vital for efficient gradient descent. A good initialization helps architecture's gradient descent to avert from vanishing/exploding problem. We would want the initialized weights to be as close to terminal weights as possible. Furthermore, real world processes usually follow normal distribution. Intuitively, approximating their terminal weights from a uniform distribution will be faster than a random normal distribution. Thus, widely used in almost all deep learning based packages, "glorot uniform" distribution is used for this experiment (Glorot and Bengio 2010).

On the other hand, optimizer for back-propagation assist the gradient descent in reaching local-minima faster. There exists a vast literature on optimal choice of optimizer algorithm for gradient descent in neural networks. Mathematical details of these algorithms are out of the scope of this paper. For initial testing of LSTM units, most popular ones were compared and the best performing algorithm was found to be "adam". Architecture of this algorithm can be found in Kingma and Ba (2015).

5 Competing Algorithms

For this experiment, FFNN and GBR were also used for the prediction task. FFNN, as has been explained in previous sections, have a feed-forward loop with single activation function in each hidden cells for approximating target values. Whereas, GBR algorithm in a homogeneous ensemble algorithm of decision trees are built out of regression. Being thoroughly researched, and popular in ML based prediction tasks, working details of both the algorithms can be found in various literature. One can refer to James et al. (2000) as guide.

6 Experimental Setup

As explained in section 3, in 2014, American Meteorological Society (AMS) hosted a competition on Kaggle for day ahead prediction of solar energy production throughout the state of Oklahoma. Daily energy production data was provided for total of 98 stations which were spread throughout Oklahoma in an irregular fashion. The task constituted of forecasting day-ahead energy based on day-ahead prediction of atmospheric variables using NWP based global forecasts (see section 2.3.1 for definition). Figure 6 shows the stations locations on a map of Oklahoma. The competition data is available for public access and researchers are allowed to submit their forecast results for comparison purposes. The large variance in topography and climate spread over the whole state of Oklahoma made it an ideal candidate for testing efficiency of learning methods. I performed initial testing of LSTM algorithm on this dataset. The major uplift in prediction score was obtained when models were built on the whole region rather than a single location. After devising the optimal working model state in order to generalize LSTM's results, I extrapolated the experiment to other locations on earth. The rest of the section describes various components of this experiment with a reference to testing procedure performed on AMS competition data.

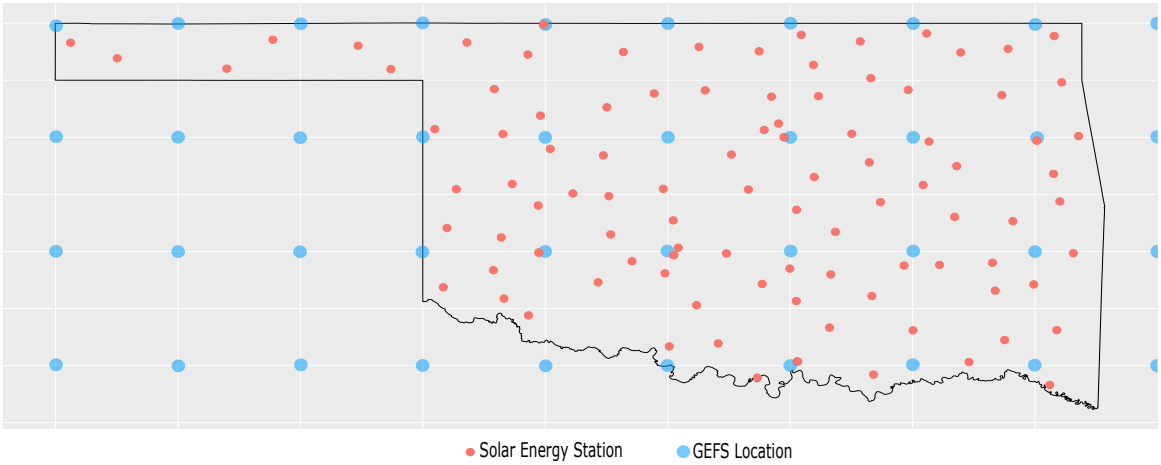


Figure 6: Geo-Spread of stations across Oklahoma in AMS competition data

6.1 Input

For this experiment no ground measurements were used as input. Table 2 summarizes the distinct variables provided to learning algorithms. Details of the nature of inputs are provided in next two sections.

Variable	Description	Units
x_dist	Haversine distance between GEFS (x) and station location	km
x_elev_diff	Difference between elevation of GEFS (x) and station location	km
day	Day of the year	
apcp_sfc	3-Hour accumulated precipitation at the surface	kg m^2
dlwrf_sfc	Downward long-wave radiative flux average at the surface	W m^2
dswrf_sfc	Downward short-wave radiative flux average at the surface	W m^2
pres_msl	Air pressure at mean sea level	Pa
pwat_eatm	Precipitable Water over the entire depth of the atmosphere	kg m^2
spfh_2m	Specific Humidity at 2 m above ground	kg kg-1
tcdc_eatm	Total cloud cover over the entire depth of the atmosphere	%
tcolc_eatm	Total column-integrated condensate over the entire atmos.	kg m^2
tmax_2m	Maximum Temperature over the past 3 hours at 2 m above the ground	K
tmin_2m	Mininmum Temperature over the past 3 hours at 2 m above the ground	K
tmp_2m	Current temperature at 2 m above the ground	K
ulwrf_sfc	Upward long-wave radiation at the surface	W m^2
ulwrf_tatm	Upward long-wave radiation at the top of the atmosphere	W m^2
uswrf_sfc	Upward short-wave radiation at the surface	W m^2

Table 2: List of input variables

6.1.1 GEFS Data

Most of the meso-scale NWP models (definition in section 2.3.2) are built upon Global Forecast models (Lorenz and Heinemann 2012). In this experiment also, 2nd generation NOAA's Global Ensemble Forecast System (GEFS) Re-forecast weather data was used as input for predicting solar irradiation. The forecast data allows a precision of $1° \times 1°$ latitude and longitude of the whole planet. The model predicts various atmospheric variables (including solar irradiation) using physical modeling techniques (definition in section 2.3.1). The prediction is made for a timestep of three hours till following three days and, thereafter, a timestep of six hours to further five days. The model is run at 00:00 UTC every day. For each atmospheric variable 11 ensemble members are created initializing with small perturbations[6]. Out of all the atmospheric variable forecasts provided by GEFS service, a few are used as inputs for the model. A list of these variables and respective descriptions can be found in table 2. Considering that the interaction of atmospheric variables along with its own trend could be highly relevant for day-ahead irradiation prediction, seven timesteps i.e from 6 to 24 hours (1 at every 3rd hour) are used. For AMS competition only five timesteps (from 12 to 24 hours) were provided.

The blue dots in figure 6 are points for which GEFS dataset is available. The task here was to create meso-scale model using information of these dots near the solar stations and minimize overall prediction error. A study performed by Aler et al. (2015) in regards with feature selection on AMS competition dataset showed that there is very little improvement in prediction score when more than four GEFS points surrounding a solar

[6] The dataset is available for public access, instructions can be found in Hamill et al. (2013)

energy station is considered. Also, from summary of results of competition, provided by McGovern et al. (2015), considering four corners of a square shaped grid surrounding the solar station was optimal choice (figure 7). The square, hence is shaped of $1^\circ \times 1^\circ$ latitude and longitude. For this experiment also, data of these four corners were used per solar station (location in focus) assuming them to be the most relevant ones whose atmospheric variable values could enhance the model's prediction. The 11 ensemble members were averaged to give one single value for one timestep per variable. While testing, mean values as input provided the best results on Oklahoma dataset for all algorithms as against mode or median values. As an alternative, one could build 11 different models for each of these 11 members and perform ensemble selection (as was done by winner of AMS competition). This procedure was avoided in this experiment primarily to save computational resources. Also, no feature engineering was performed using these variables. From table 2, except the first three variables, all inputs are GEFS sourced atmospheric data. So, as final inputs, for all the locations tested, each of these variables had seven values i.e one, mean of 11 ensembles, at each timestep.

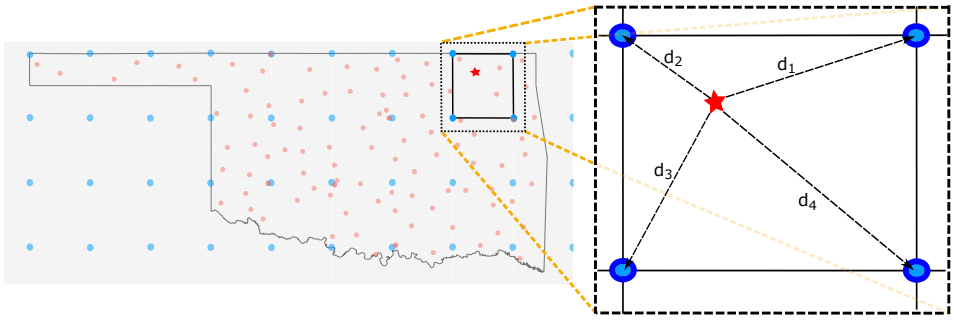


Figure 7: A schematic view of single grid surrounding the location

6.1.2 Spatial/Temporal Data

Apart from atmospheric variable values on the corners of surrounding square for a relative measure and helping model to learn the spatial structure, some topographical variables were also provided as input. The first 3 input variables from table 2 are geo/temporal based inputs. The difference between the elevations of the GEFS corner point and location in focus was given to the model. Elevation of the location in focus was downloaded from Google's API service[7]. The distance between location and GEFS corner was calculated using Haversine Distance formula. Figure 7 shows an isolated location that needs to be forecasted. Finally, the day of the year was also given as an input. Hence, for each station's corresponding variable values against four corners in the grid were provided as input.

[7]Python based "geocoder" module was used for gathering elevation data from Google Maps service.

6.2 Target Variable

Solar energy data requires a PV setup, unfortunately, solar energy companies do not provide it for public access and datasets similar to that of AMS competition is rather rare to obtain. An alternative method is to predict Global Horizontal Irradiation (GHI) instead. GHI is also referred to as short-wave downward radiation or global radiation in the literature. Researchers usually use this value as surrogate PV energy values. GHI to PV conversion procedure has been explained in Lorenz et al. (2010). The common way of recording GHI is via installation of Pyranometers on a planar surface. However, there are two major issues which arises from using ground based GHI readings.

First, installation of these instruments is a resource consuming operation. Consequently, the installed apparatus requires constant monitoring and repairing. Therefore, the GHI readings are available only for specific locations around the globe where they have been installed. As explained in section 3, all researchers have built models on locally acquired data through either meteorological stations or research fundings. This local data pertains to specific temporal and spatial properties making it biased towards a limited regional and climatic conditions. Hence, such procedure hinders in showing model efficiency on a global scale.

On the other hand, pyranometers require constant calibration for capturing correct GHI values. Due to various inherent systematic uncertainties which are propagated, the final readings have a relatively large degree of uncertainty. Models based on such values cannot be validated or verified to a level of accuracy higher than that of the measurements. When compared against physical equation outputs, these readings have shown to have $\approx \pm 5\%$ error (Sengupta et al. 2015, Section 3.3). Apart from which, other quality problems often cause unexpected large GHI readings at night and small readings in day duration. Correction of these readings to real values should ideally be done by the local station manager, which often is not possible when working with retro-data.

As a proxy target variable GHI readings "now-casted" via satellites are used in this experiment. As explained in section 2.2, geo-stationary satellites based solar models uses cloud images in real time to predict their motion and give a very short-term forecasted GHI. US based National Solar Radiation Database (NSRDB) allows free access to Physical Solar Model (PSM) and Europe's Copernicus Atmosphere Monitoring Service (CAMS) to its Radiation Service[8]. Available for the period of 1998-2014, this dataset allows a precision of 4×4 km with a temporal resolution of up to 30 minutes for the US and a few countries in neighbouring continents. When checked against the AMS competition's target data of each station I found a correlation of 0.95 to 0.99 for every year during 1998-2007 pe-

[8] For NSRDB's PSM, one can refer to (Sengupta et al., 2012) for details on method involved and validation of calculated GHI values against real GHI

riod. As for CAMS Radiation service, which allows similar spatial resolution, and down to 1 minute of temporal resolution, corresponding correlation values are claimed to be between 0.91 to 0.97 in (Schroedter-Homscheidt, 2016), and is available for countries in Europe, Africa and a few other countries in South America and Middle East.

Even though there is inherent error in these satellite based GHI values, a major advantage over ground based GHI measurements is that they are available for every location (of specified resolution) in the aforementioned locations. Additionally, there are no missing values in the data. The inherent error would be present in both ground and satellite based GHI values but the availability of satellite data makes it useful for multi-location validation. Apart from recommendations from domain researchers (Lorenz et al. 2009b; Lorenz and Heinemann 2012; Tadesse et al. 2016; Sengupta et al. 2015), the correlation values found against Oklahoma's solar energy dataset also advocated for using satellite values as proxy target variable.

6.3 Region Construction

From the results obtained via initial testing on the AMS competition dataset, it was clear that regional modeling outperforms location specific modeling. Taking advantage of vast and detailed spatial availability of target variable, as explained above, a learnable region within the small $1^\circ \times 1^\circ$ grid was designed. Inspired from the nature of AMS competition dataset and initial results, fake solar energy stations were built inside the grid. The underlying idea was to create a grid level model where algorithm learns spatial and temporal behavior of the region pertaining to irradiance received in it. Figure 8 shows a region thus created with fake stations. A total of 49 fake stations were created with 36 uniformly distributed (at 0.2° from horizontally or vertically placed neighboring fake station) and 13 others were created randomly. With the given resolution of satellite data, it was safe to build the uniformly distributed fake stations at such intervals so as to avoid introducing bias. Including station for which forecast is required, algorithms had 50 stations to learn from. With ease in replication and implementation, this input architecture allows analysing over any part of the world where high precision satellite based GHI values are recorded. Ergo comparative studies of efficiency of prediction algorithms can be performed.

6.4 Validation

From the nature of solar irradiation forecasting using global-scale models, dependencies on location specific climatic conditions and geo-spatial spread are conspicuous. An objec-

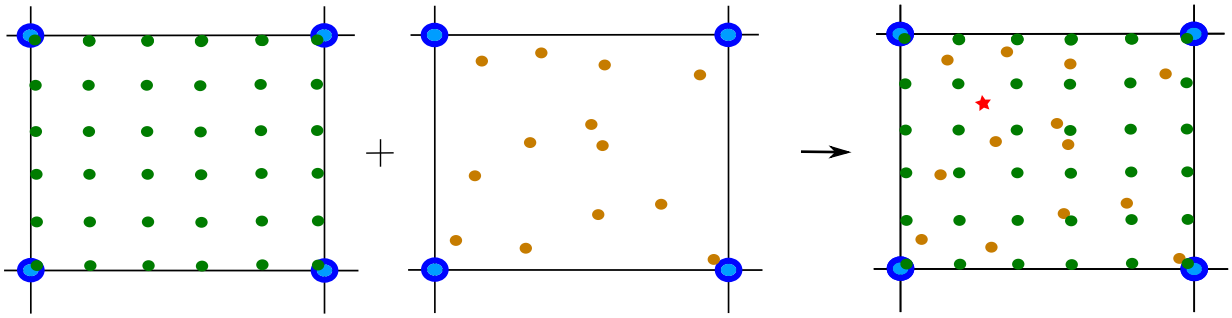
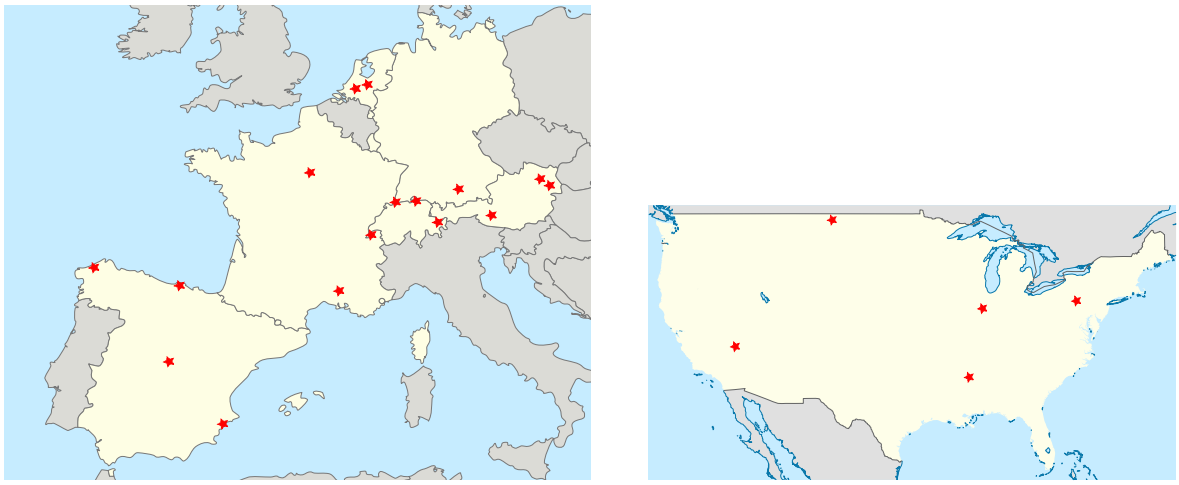


Figure 8: Building of fake stations- 36 uniformly and 13 randomly distributed. Star represents station in focus.

tive validation of comparison of efficiency of learning algorithms requires multi-location based experiments. Diverse set of climatic conditions and topographical properties of chosen locations would help in judging the robustness of learners. For this experiment 5 locations were chosen from United States of America and 16 from 6 different countries of mainland Europe. Table 3 contains information on all the selected locations in focus. Also, it can be seen through figures 9a and 9b, the locations are chosen pragmatically so as to maximize diversity. For each of these locations, the region (section 6.3) was constructed to perform the experiment.



(a) Europe based locations

(b) USA based locations

Figure 9: Locations on Map

Data was collected from the year 2005 to 2014 which was further divided into training, validation and test datasets. The test dataset is hold-out dataset, which is never touched until the final models are built. For this experiment data pertaining to year 2014 was used as test dataset. Validation set, year 2012 and 2013, decided model parameters (more on this in section 6.6). The rest of the data, i.e from 2005 to 2011, was used for training the models. The design of this experiment was inspired from the initial testing on AMS competition data. The competition data constituted of 14 years (1994 to 2007) of daily

Country/Location	Latitude	Longitude	Elevation (in m)	Climate
Germany				
Hohenpeissenberg	47.8	11.01	925	Continental
Netherlands				
Cabauw	51.97	4.93	-1	Humid continental
De Bilt	52.1	5.18	2	Humid continental
Austria				
Grossenzersdorf	48.2	16.57	154	Continental
Sonnblick	47.05	12.96	3013	Alpine
Wien	48.25	16.35	215	Continental
France				
Carpentras	44.08	5.06	100	Mediterranean
Palaiseau	48.71	2.21	158	Continental
Spain				
Santander	43.48	-3.8	69	Oceanic
La Coruna	43.35	-8.42	43	Oceanic
Alicante	38.28	-0.55	18	Mediterranean/Humid
Madrid	40.45	-3.73	634	Continental
Switzerland				
Basel	47.55	7.58	273	Temperate Atlantic
Zurich	47.48	8.53	427	Moderate maritime/continental
Geneva	46.25	6.13	419	Moderate maritime/continental
Davos	46.82	9.85	1559	Continental/Alpine
USA				
Desert Rock	36.63	-116.02	1007	Arid
Bondville	40.07	-88.37	213	Continental
Fort Peck	48.32	-105.10	634	Continental
Goodwin Creek	34.25	-89.87	98	Humid Continental
Penn State	40.72	-77.93	374	Humid Continental

Table 3: Locations and their climate types

solar energy readings of 98 stations across Oklahoma region (figure 6) which in itself was sufficient to train, test and validate the models. Data pertaining to years 1994 to 2002 was used for training, and succeeding years for validation (first 2 years) and test (last 2 years).

6.5 Data Pre-Processing

6.5.1 Handling Missing Values/Outliers

The major reason for pre-processing the data is because of the fact that real world datasets are never clean or model-ready. They often contain missing values and outliers which, in most of the cases, are a result of human or machine error. A judicious approach is required to find these mishaps and normalize them. However, these are also one of the most controversial and debated topics in the field of predictive statistics. It has been argued that both missing values and outliers are information in themselves and one should refrain from imputing them. Wainer (1976) also introduced the concept of the "fringelier",

referring to "unusual events which occur more often than seldom" (p. 286). The techniques of imputation are also often debated as scientists have different point of views. Mean, median, weighted mean or periodical average are some of the values used and there is never one particular solution that fits in every situation. In the ideal scenario, the data collectors/managers should be the one handling these issues as they have the field experience and know the reasons for every missing or outlier values. However, such scrutiny is often not possible.

On the account of aforementioned reasoning and due to limited domain knowledge, I refrained from performing outlier analyses for this experiment. On the other hand, GEFS data had ~0.3% and ~0.5% missing values for locations based in the US and Europe respectively. Assuming that the missing values of atmospheric variable values would most definitely be an error at data capture/storage process, following steps were taken to correct them:

- For each distinct GEFS based atmospheric variable (ref. table 2), month-wise data was separated. (i.e 12 sub-datasets with for each variable with data from year 2005 to 2014)
- For every timestep (7 in total) of each variable, it was checked if any missing values exist.
- A linear regression fit decided value of missing data of the respective timestep. The linear model was built using all other timestep values as predictors. So, for eg. if data of a variable for timestep of 12 Hr is missing then a fit trend based on other timestep values (6,9,15,18,21,24) of that variable in specific month would impute the missing value.

6.5.2 Input Structure

For LSTMs, due to the specific learning behaviour of sequential data (see section 4.3.1), modifications were made to input structure for better comprehensibility. Figure 10 represents the schematic view of this structure. The temporal nature of atmospheric variables was explained via ordering timesteps in their sequential format. In the figure, for per-day observations of a station, the ordered sequence of all variables was provided to LSTM units. For intelligibility, the blue line shows distribution of interaction of all the variables throughout the day with red dots representing values(all variables combined[9]) forecasted at each timestep. Information pertaining to time-series behaviour can be effectively extracted using this architecture. Interaction of atmospheric variables at each timestep was undertaken through inner activation functions of LSTM units. The non-temporal based

[9] Please note that this combination will happen inside the LSTM unit through activation functions

variables (first three in table table 2) were given with each timestep. While testing on AMS competition data, as a hypothesis of capturing solar energy's dependency on longer ranges of past data, this ordered sequence of inputs was extended to previous 1 to 30 day's observations. However, with Occam's razor in play, best prediction scores were obtained via only 1 day (respective present day) forecasts lookup. Furthermore, inter-day and inter-station data was shuffled while feeding the data to LSTM units.

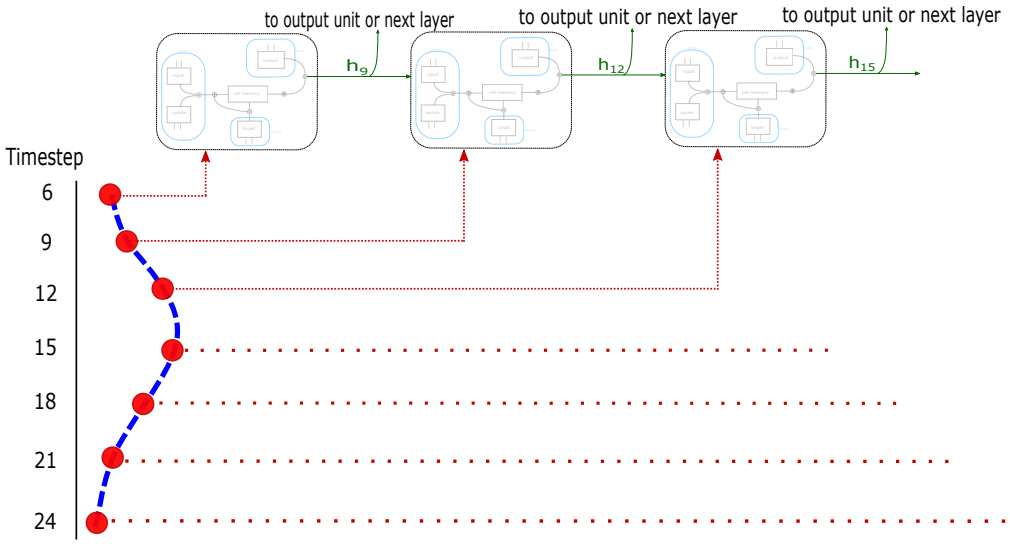


Figure 10: Time ordered structure of input for feeding into LSTM units

On the other hand, for inputs of other algorithms traditional method was adopted, i.e. all timestep values of each variable were provided in a single row against per-day observation of respective station. This input structure also was used by winners of the AMS competition.

6.6 Model Tuning

From the results of the AMS competition(McGovern et al. 2015), Gradient Boosting Regression(GBR), showed an upper-hand in predicting solar energy against most of state of the art learning methods including FFNN. Also, in research, the most popular algorithms used in GHI forecasting is FFNN (see table 1). The two learning algorithms were, thus, compared against LSTM. As an additional measure Persistence, which has been often used as benchmark in relevant literatures for comparison, is checked. It can be understood as assuming a scenario where there is no change in consecutive periodical measurements of readings. Hence, the Persistence model projects the measured target value at $t - 1$ to time t as prediction for time t.In our case for day ahead prediction of GHI values t is 24 hours or one day and can be expressed via (12).

$$GHI_{persistence,t} = GHI_{measured,t-1} \tag{12}$$

The initial testing performed on AMS competition data with various LSTM configuration assisted in discovery of optimal model skeleton. As have been mentioned by Hand (2006), in most of the comparative studies, researchers tend to over-tune the models in order to achieve better performances against comparing algorithms. Keeping this in mind, multiple locations were considered for checking efficiency of each algorithm. Other than that, the best performing model structure of LSTM was further used. This structure, or the penultimate model skeleton, contained two layers of LSTM units followed by a drop-out layer and finally, a squashing sigmoid layer.

6.6.1 Transfer Learning

An important hypothesis of this experiment was that the learning algorithms are approximating or learning the whole region (sec. 6.3). Built upon similar hypothesis, traditional methods of mesoscale NWPs generally take weighted/pixel averages of forecasted variables based on the distance of location in focus from GEFS points. I have refrained from tuning model parameters for every single region. Instead, only one region in the US i.e. Penn State, was used for model tuning. Penn State had the highest intermediate variability of GHI values Persistence model was used to check this measure. Assuming the location to be hardest to learn, method of grid-search of optimal parametric values was used to obtain best model configurations. The values used for meta-parameters can be found in appendix in table 10. Using the optimal parameter values, thus found, models were built of each learning algorithm at all the locations. Thus, using the concept of *Transfer Learning* which theorizes the applicability of already gained knowledge of algorithms in similar tasks without having to learn from scratch. One parameter was still set loose to give algorithms some degrees of freedom. For LSTM and FFNN, it was the number of iterations and for GBR it was the number of trees. A pool of models was finally created for each learner. The model providing least error score on whole region for validation dataset was chosen to be the final model. The error metrics are explained in the next section.

6.7 Performance Metrics

The performance metrics chosen for comparing the prediction accuracy of algorithms are Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). RMSE accounts

for sensitivity for large errors resulting in higher costs for users whereas, MAE is useful where energy costs are proportional to forecast error. In AMS competition, prediction accuracy was checked via overall MAE of station energy forecasts of the whole region (Oklahoma). RMSE and MAE are calculated by equations (13) and (14). N represents the number of observations and I_x represents the value of target variable with $x \in (predicted, measured)$.

$$RMSE = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^{N} (I_{(pred,i)} - I_{(meas,i)})} \tag{13}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |(I_{(pred,i)} - I_{(meas,i)}| \tag{14}$$

The AMS competition only used MAE as performance metric. Hence, while initial testing only MAE was used to obtain the optimal LSTM model skeleton.

6.8 Programing Details

The weather based GEFS data can be downloaded in netCDF4 format. R programing (3.3.1) was used for initial extractions and pre-processing of this data. Python (2.7) was used for LSTM and FFNN modeling via Keras (Chollet 2015) repository of DL. Both Python and R have been used interchangeably for various tasks. Initial parts of the code were run on Windows 8 machine with 4 cores and the rest on Ubuntu based Linux system with 48 cores. It is important to mention because parallel programing using multiple cores was extensively used in the experiment and one would need to consider this while using these codes. The final source codes of this comparative experiment can be found at https://github.com/atshikharsrivastava/SolarEnergy-Pred

7 Results

7.1 Initial Testing

From the initial testing performed on the AMS competition dataset, the robustness of LSTM algorithm became apparent. GBR, the winning algorithm of the competition, was used for comparison against the LSTM model during initial testing. Since the competition used only MAE as performance metric, RMSE was not calculated for this testing. Table 4 shows the comparison of final MAE scores of LSTM model against GBR. Two regimes of modelings were performed: first was the per-station modeling, the final MAE was aggregate of MAE of model built on each station and the second was regional-modeling, where just one model was built for whole Oklahoma region. The final LSTM model skeleton, discovered from initial testing, tuned over the validation set for optimal number of LSTM neurons is used for comparison here. Whereas, GBR's base configuration were used from top scoring models from the AMS competition with an exception of number of trees which was optimized using validation set.

Type	**GBR**	**LSTM**
Per-Station Modeling	2.0144	**1.9903**
Regional Modeling	1.9373	**1.9145**

Table 4: MAE of predicted energy for AMS competition dataset (unit: MJ m^{-2})

Furthermore, the competition host site, Kaggle, still allows entries to check model performances against top scores of the competition. Using the whole data (1994-2007) for training, with inputs same as mentioned in section 6.5.2, and implementing a 3-fold cross validation, the resulting LSTM model outperforms the best model of the competition. These results served as the initial motivation for generalizing this result on a larger scale.

7.2 Experiment on other locations

For generalization of results from testing over AMS competition data, various locations across the planet were used for experimenting. Table 5 provides RMSE and MAE values for experiment performed on all the selected locations. For almost every location, superiority in GHI prediction of LSTM is shown.

Table 5: RMSE and MAE of predicted GHI for chosen locations (unit: W m^2)

	RMSE				MAE			
Location	**FFNN**	**GBR**	**LSTM**	**Pers**	**FFNN**	**GBR**	**LSTM**	**Pers**
Germany								
Hohenpeissenberg	30.03	29.79	**29.26**	65.48	21.28	20.91	**19.71**	45.69
Netherlands								
Cabauw	28.06	28.05	**25.42**	54.20	20.00	19.25	**18.67**	38.09
De Bilt	24.96	23.81	**23.76**	53.22	17.30	**16.66**	16.91	37.15
Austria								
Grossenzersdorf	**26.84**	30.67	27.04	61.19	20.24	22.20	**19.57**	42.47
Sonnblick	36.27	**34.89**	36.98	62.50	26.33	**25.39**	25.81	44.46
Wien	31.59	34.07	**27.94**	62.99	23.47	24.71	**20.34**	43.74
France								
Carpentras	26.87	25.70	**23.90**	63.05	18.85	18.71	**17.31**	42.11
Palaiseau	29.30	**26.46**	31.38	56.89	19.69	**19.00**	19.79	38.70
Spain								
Santander	31.43	31.93	**30.34**	67.95	22.45	22.68	**21.53**	49.20
La Coruna	33.34	31.21	**30.92**	59.50	24.18	22.24	**21.23**	42.87
Alicante	27.16	27.47	**26.98**	50.85	18.79	20.88	**18.11**	34.31
Madrid	25.43	**22.83**	23.60	50.31	19.12	16.22	**15.67**	33.02
Switzerland								
Basel	28.59	27.63	**24.99**	59.45	20.45	19.92	**17.95**	40.33
Zurich	30.61	28.46	**27.97**	61.52	23.00	20.70	**20.12**	40.96
Geneva	28.95	28.12	**28.07**	60.82	20.87	20.43	**19.64**	41.52
Davos	**32.12**	32.20	32.69	67.19	24.13	24.44	**23.78**	48.24
USA								
Desert Rock	29.38	27.29	**25.77**	47.07	17.58	**14.97**	15.24	29.63
Bondville	43.85	39.67	**37.78**	74.86	37.25	28.80	**27.94**	53.90
Fort Peck	37.00	35.82	**33.49**	59.36	28.18	25.81	**25.08**	43.50
Goodwin Creek	36.47	31.76	**30.89**	73.39	23.28	22.62	**22.49**	55.20
Penn State	38.71	36.39	**35.02**	77.48	29.53	26.31	**25.01**	57.65

Bold faces represent best prediction score.

Pers abbreviates Persistence model.

8 Discussion

8.1 Drawbacks

LSTMs proved to be better learning algorithm than GBR or FFNN. All three of them have showed high accuracy against the benchmark persistence model. However, there is no free lunch. ML based predictive modeling heavily relies on optimal combination of meta-parameters of the base algorithm. Finding out the specific combination depends on the nature of problem, dataset, and selected base learners. This, in turn, makes the optimal values swerve and difficult to find. Consequently, it requires a trial and error method and, hence, increased resource/time consumption. The larger the number of meta-parameters the harder it gets to locate the optimal choices. Especially for LSTMs, the number of layers, number of units, choice of activation functions at each stage, subsequent choice regularizations and their values, values for weight initialization, choices of optimizers and their inner configuration are the basic meta-parameters to adjust. This increases the search space exponentially and thus, requires arduous judgments and testings for finding optimal configuration. The large degree of search space can be used to argue that model configuration used in the experiment might not even be the best choice. Further exhaustive testings with permutations and combinations of parametric choices will be required. Similarly, GBR and FFNNs have large meta-parameter search space as well. However, in case of GBR, configurations near to the winning model of AMS competition have been used on the assumption that they are proven to show better accuracies. The other drawback is also a general problem of the ML world. LSTM method, like many others, is a black-box model, i.e. finding out the exact intricate dependencies of variables on each other is not possible.

The experiment design, on the other hand, is built upon GEFS data to predict Satellite based GHI values. The design can only be replicated for places on earth where both GEFS and Satellite data are available. For most of the locations this data is attainable (see Sengupta et al. (2015) for available databases).

8.2 GEFS Data

One important thing to note here is that the GEFS weather forecast variables are used from a free service, the NOAA's GEFS reforecast2. It allows a precision of $1^\circ \times 1^\circ$ latitude and longitude. This experiment is based on the premise of accuracy of these forecast variables. There are better resolutions available. For example, Europe's ECMWF IFS gives a resolution of $0.125^\circ \times 0.125^\circ$, however, it is not available for public access. Thus, the GHI prediction results obtained cannot be classified as state of the art scores. With

increase in resolution, the experiment can be further replicated, tested, and used to get the best results. From results obtained in this experiment, it can be safely said that LSTMs could provide higher accuracy than the existing methods when provided with the best resolution available.

8.3 Tests Performed

During the initial testings on AMS competition data, various input structures were used. When using the time-series style input (see section 6.5.2), the input sequence was extended than only using forecast of 1 day (which is to be predicted). The extension was tested for the previous 1-30 days with the hypothesis that there could be a trend that the algorithm would capture. However, the results suggested otherwise, only day-ahead forecast data was found to be most predictive. Regarding selection of inner activation functions *sigmoid*, *tanh* and *ReLu* were tested. *ReLu*, here, proved to be ideal for forecasting solar energy via learning on atmospheric variables, figure 14 in appendix shows the general behavior over training and testing on AMS competition dataset. *ReLu* has also proven to be successful in other application domains of neural networks such as acoustic learning or Boltzman machines (Maas et al. 2013; Nair and Hinton 2010). In addition, the spatial averaging of GEFS variables using distance from grid corners as weights was used. Initially, when only single layered LSTM models were built, this method worked better as it included spatial information by usage of distance as weights. The method also saved computational resources to a large extent as number of input variables were decreased by a factor of 4. However, this changed when a two layered model was considered. Two layered LSTM models approximated target values much better and provided the model skeleton for this experiment. Solving of higher order dependencies/relationships was, thus, in action to learn the spatial nature of region. A three layered model was also tested which couldn't perform better than the two layered model. Further addition of layers were, hence, not checked.

A major improvement in prediction scores was observed when instead of building models for each station, one single model for the entire Oklahoma region was built. This is also reflected in table 4. The result proves that regional modeling is significantly better than per-station or location based modeling. Location based modeling limits training data and induces bias. This result also confirms a study performed by Lorenz et al. (2009b).

8.4 AMS Competition

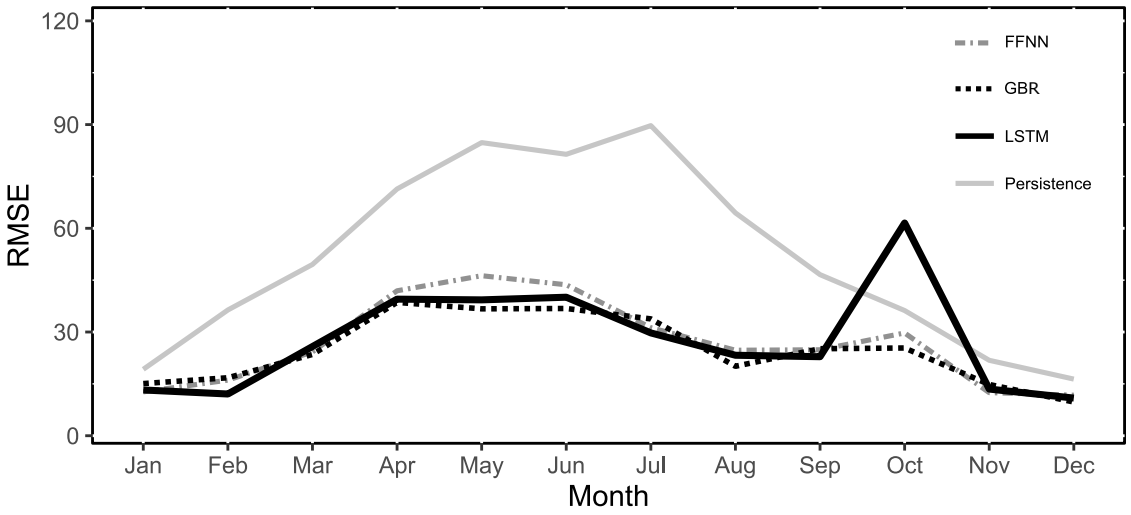
The winners of the competition used exhaustive machine learning methods. The top scorer created 13 different base models for each ensemble member with 3 fold cross

validation. Furthermore, the team also used an ensemble selection method to pick and combine the best performing candidate. The second ranker used extensive feature engineering methods for increasing the model's predicting capabilities. And, the third ranker encompassed time-series behavior via feature engineering of top predictor. A linear fit of all timesteps of this variable was included in input, which can be safely said to be specific for Oklahoma's weather behavior (McGovern et al. 2015). The LSTMs were able to render higher accuracy with raw inputs in almost all the tests when compared against GBR. When provided with input structure used by third ranker, LSTMs also accessed the gain in knowledge of engineered variable and still outperformed GBR. On the other hand, during local testing period of this experiment no weather based feature engineering was performed. As mentioned earlier, when trained with all the data provided in competition, LSTMs achieved a higher score than top scores in the leader-board. Even though multiple tests had to be done to reach the optimal model structure, which could be suggestive of over-tuning, the simplicity of final structure and consistent better performance inspired to extend the experiment on other locations around the world.

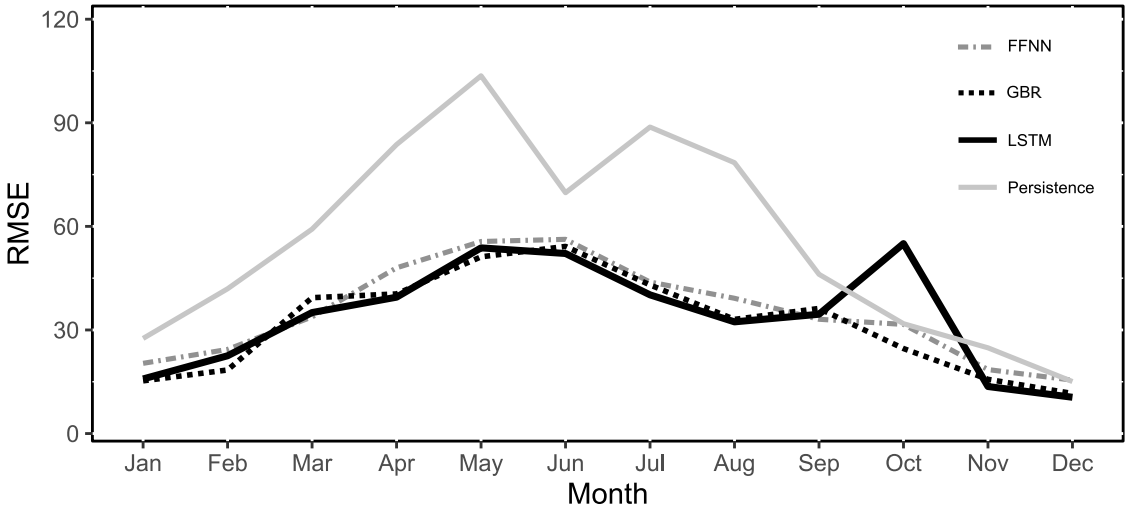
8.5 Performance on Experimental Datasets

The results of this experiment, as shown in table 5, comprehensively proves efficiency of LSTMs for GHI prediction task in hybrid modeling regime. There is minor fluctuation in best models when MAE results are compared with RMSE results. From the degree of variation these fluctuations can be neglected. Except two locations, Palaiseau and Sonnblick , where the LSTM models showed relatively poor performance, the performance at almost every locations is better. Interestingly, a month wise analysis of RMSE of these locations, as can be seen in plots 11a and 11b, shows LSTM's unexpected performance deviation for October month. Further analysis showed that only for October month LSTM performed significantly worse than GBM and FFNN models for almost every location in Europe. In October, the average loss in RMSE using LSTM over GBR and ANN was 34.5% and 26.6% respectively. On the other hand, this value for almost every other month shows LSTM's primacy over these two algorithms. Table 6 shows the RMSE and MAE values of test data when October month is not considered for Europe based locations. LSTM proved to be better for every location in this scenario. However, it should be noted that LSTM's October performance is not worse for every single location. For locations such as Basel, Zurich or Carpentras the performance drop was observed by all three algorithms, GBR and FFNN predicted worse than LSTM. A deeper look, like variation in day-wise prediction errors, only pointed towards the conclusion of October being a difficult month to generalize for locations in Europe, more for LSTM than others.

It could be criticized that this analysis is highly LSTM centric. However, looking at the



(a) Palaiseau



(b) Sonnblick

Figure 11: Month-wise RMSE plots for worst performing locations of LSTM

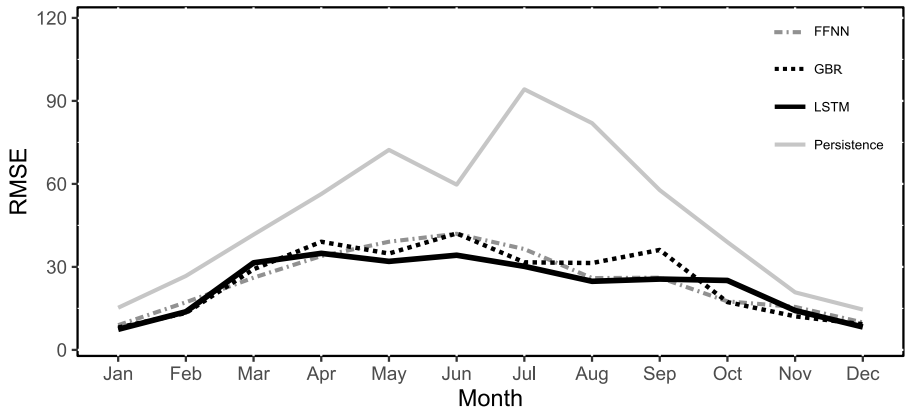
Table 6: October Removed - RMSE and MAE of Europe locations

	RMSE				MAE			
Location	**FFNN**	**GBR**	**LSTM**	**Pers**	**FFNN**	**GBR**	**LSTM**	**Pers**
Germany								
Hohenpeissenberg	29.89	28.88	**28.19**	67.41	21.13	20.26	**18.99**	46.02
Netherlands								
Cabauw	28.76	28.84	**25.45**	60.82	20.62	19.76	**18.71**	41.61
De Bilt	25.72	24.08	**23.34**	60.09	17.85	17.68	**16.71**	41.57
Austria								
Grossenzersdorf	26.30	29.12	**25.49**	60.29	19.74	20.96	**18.52**	41.25
Sonnblick	36.72	35.69	**34.84**	57.94	26.70	25.90	**25.67**	41.30
Wien	31.38	32.71	**25.65**	59.33	23.07	23.45	**18.92**	40.32
France								
Carpentras	24.73	25.45	**23.05**	63.23	17.72	18.44	**16.62**	40.64
Palaiseau	29.32	27.56	**26.92**	58.30	19.62	18.97	**18.48**	39.85
Spain								
Santander	31.98	32.81	**30.36**	73.93	22.92	23.46	**21.77**	50.86
La Coruna	33.10	31.26	**30.71**	66.52	23.92	22.22	**21.10**	46.42
Alicante	27.07	27.69	**26.44**	52.50	18.83	20.68	**17.82**	34.26
Madrid	24.89	23.02	**21.61**	52.65	18.68	16.30	**14.94**	35.06
Switzerland								
Basel	28.77	27.98	**25.15**	60.93	20.38	19.91	**17.83**	43.83
Zurich	30.05	28.63	**28.26**	66.04	22.70	20.63	**20.23**	45.71
Geneva	28.13	28.01	**26.95**	64.43	20.95	20.31	**19.64**	44.02
Davos	32.05	32.17	**30.83**	61.37	23.90	24.22	**22.79**	44.11

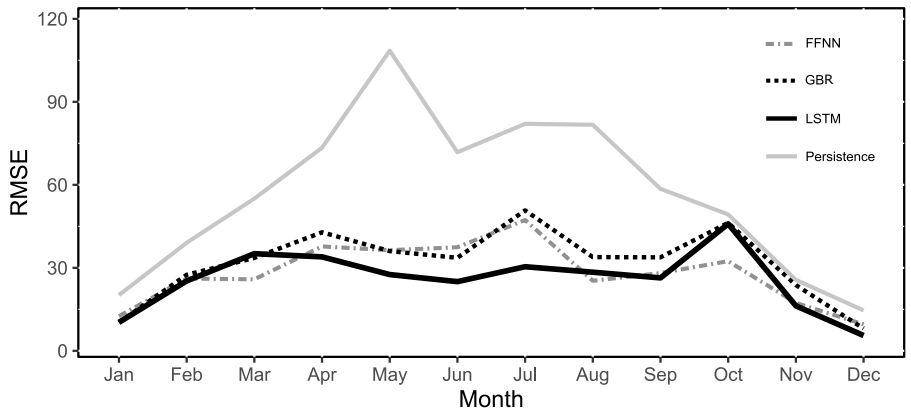
Bold faces represent best prediction score.

Pers abbreviates Persistence model.

month-wise plots of Wien and Cabauw (figures 12a and 12b), where LSTM is significantly out-performing GBR and FFNN, it is apparent that both algorithms are consistently performing poor for almost all months. It can be also seen from tables 7 and 8 in appendix where the overall and monthly improvement in RMSE scores using LSTM over GBR and FFNN are calculated. On the other hand, only little of this unexpected performance drop, like October in Europe case, was observed in locations chosen from the US. Plots 13a and 13b provide month wise RMSE values of relative best and worst performing locations for LSTM which are Fort Peck and Goodwin Creek respectively. Given the industry-wide usage of RMSE as performance indicator, this analysis was not performed on MAE as thoroughly. Similar results are expected given its parallel in final results with RMSE values. Application of ensemble selection over these models are expected to increase the prediction accuracy. However, this experiment was solely concerned with comparison of base algorithms.

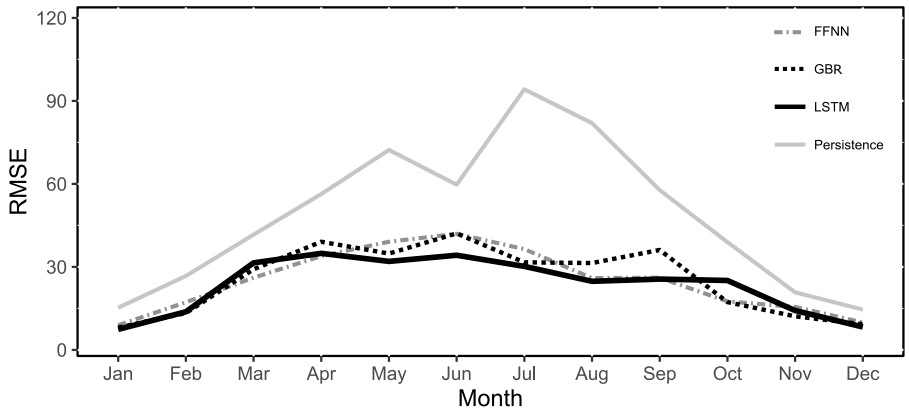


(a) Cabauw

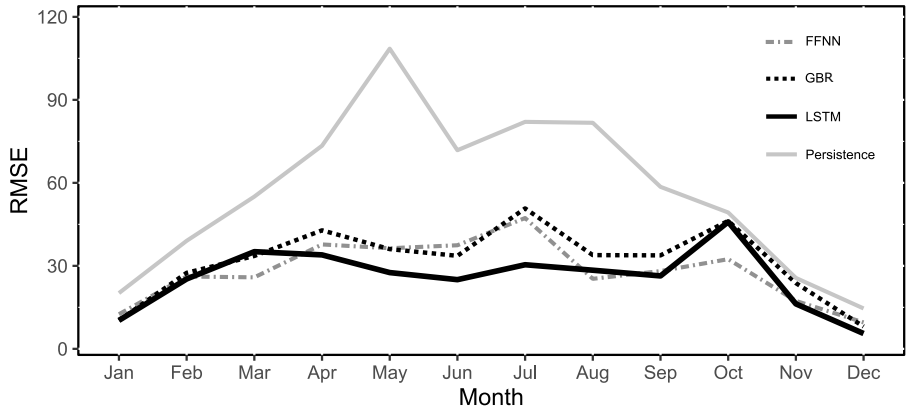


(b) Wien

Figure 12: Month-wise RMSE plots for best performing locations of LSTM



(a) Fort Peck



(b) Goodwin Creek

Figure 13: Month-wise RMSE for best (a) and worst (b) LSTM performance locations in US

8.6 Resource Assessment

For resource assessment, planning authorities usually look at the long-range horizon availability of solar energy in the location (Sengupta et al. 2015). Even though a location might provide annually high solar energy, for PV setup, as an initial measure, the short-horizon predictability must be taken into account for efficient early participation in energy auction market, for when the respective PV system goes live. As shown in this experiment, the inherent intermittent nature of solar resources varies over a large degree over different places. Persistence score, by definition, explains day-ahead variability in GHI of a location. The correlation of RMSE scores between Persistence and other ML models is ~0.6, i.e. increase in Persistence error does not necessarily mean an increase in errors of ML algorithms. ML models have shown to understand atmospheric behavior even in cases of high-variability. For example, in graph 12b from the month of March to May the Persistense errors show increasing trend. ML algorithms, here, show opposite

trend, their accuracy increases for these months. A retrospective check using this experiment design, which allows remote modeling of any location on planet, will help related personnels in decision making regarding optimal location choices. Together with LSTM algorithm, the design of this experiment in GHI forecasting task makes it an ideal tool for such retrospective checks.

8.7 Sequential Prediction

LSTMs are primarily known for their ability of understanding deep contexts in sequential data. This served as the principal motivation for its application in GHI forecasting task. It is worth mentioning again that LSTMs have shown high efficiency in producing sequences as well (see examples of application in section 4.1). The GHI target data used in the experiment was available for a resolution of at least 30 minutes. Half-hourly based sequential output for next day (or required horizon) can be generated using this target sequence. For intra-day energy auctioning participation or for a better knowledge of day-ahead fluctuations this would provide an added advantage. The functionality has not been used in the experiment primarily due to inability of competing algorithms to do so. More importantly, GBR is computationally expensive for building models, using this methodology for building models for every half hour would have increased the computational time-frame of the experiment by an order of 24×2. Nonetheless, it should be considered for future applications.

9 Conclusion

The increasing advance of solar related technologies into the energy market is creating a demand for highly accurate solar based energy forecasts. Using a case study of New England based operators, Brancucci Martinez-Anido et al. (2016) categorically showed the significant value-add in cost savings when accuracy of day-ahead forecasting of solar energy is improved. The standard practice in solar energy forecasting of a location includes usage of meso-scale NWP models which are built upon Global forecasts of atmospheric variables. These models based on physics have evolved over the decades of research and development using super-computers. Although their ability in explaining intrinsic relationships has earned them industry-wide trust, the complexity of these models now hinders them from significant improvement in forecasting. Only a gradual growth can be expected in this field of research (Larson 2013). On the other hand, hybrid models using supervised ML algorithms are becoming increasingly popular for showing promises of gaining such accuracy. From study of related research, the lag of adaptation of new algorithms is apparent.

The experiment that I have performed provides three major contributions in GHI forecasting domain. First and foremost, this comparative study proves robustness of LSTM algorithm over GBR and FFNN for day-ahead GHI predictions. Studies regarding competing of algorithms face criticism of biased results. However, this experiment was pragmatically designed to check efficiency of these algorithms at many locations in the US and mainland Europe. Application of Transfer Learning method, which was used to extrapolate model structures, ensured unbiased and accurate prediction results for each location. This design of experiment is the second major contribution to the research field. Most ML based studies have used only a single location due to measurements related constraints. The results of those studies are, thus, location specific and need to be tested at other locations with different spatial/climatic attributes. The experiment of this paper obviates this traditional flaw by using only remote-sensing data. ML researchers in the domain can use the design of this experiment for future comparative studies of predictive algorithms and test for almost any location where highly accurate nowcasted GHI data is available. Third, this experiment design and LSTM algorithm, both could be efficiently used for remote solar resource assessment. Apart from solar resource availability of a location gathered from long range horizon models, its short-term predictability must be acknowledged for efficient market biddings in the future. As an integrative tool in decision making process, the planning authorities can use both of them in conjunction to remotely check retrospective day-ahead forecasting accuracies of any tentative location.

10 Future Work

As the use and application of technology continues to grow, there is an expected increase in accuracy and resolution level of satellite based weather data. The third generation Meteosat will see the launch of six new satellites from 2021[10]. These satellites are estimated to provide a nowcasting with resolution of 1×1 km. On the other hand, the plans of extensive PV installations, such as that of Germany (Wirth 2017), will require highly accurate GHI forecasts for effective market penetration. The work in this paper and contemporary methodologies being used in the domain are far from sufficient for this need. There are ample amount of gaps that need to be exploited using state of the art ML techniques. As a next step, LSTM model needs to be developed using Global forecasts of best resolutions available. Most of the solar energy predictions still use outputs of mesoscale NWP models. A comprehensive comparative benchmark study using state of the art NWP and Hybrid models is highly important. It is expected that LSTMs will also out-perform most of the NWP models. Furthermore, mesoscale NWPs could be used as additional inputs in LSTM models to achieve even better accuracy. The results of this experiment provide an inspiration for testing LSTMs for various forecasting horizons. Other solar based technologies like Concentrated Solar Power which is based on a similar concept, could be assisted with LSTMs efficiency. Also, the proven ability of LSTMs, and other newer RNNs like Gated Recurrent Unit for learning and predicting sequences, should be used to obtain sequential hourly or even half-hourly resolutions in day-ahead prediction. Finally, alternative DL methods like CNN or Bayesian DLs or combination of CNN and modern RNNs have been shown to produce promising results in various fields. In this domain, for example, CNNs could be used for spatial averaging using each weather variable as a filter to learn the whole region. The expanding universe of predictive modeling needs to be thoroughly integrated with the solar energy domain to advance their economic feasibility, efficient market penetration and thus, carve a path to the biggest source of energy.

[10] www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/index.html

Acknowledgement

I would like to thank my supervisor Prof. Dr. Stefan Lessmann for his guidance and support. I would also like to thank Dr. Patrick Jähnichen of Humboldt University's Mathematisch-Naturwissenschaftliche Fakultät Institut für Informatik for giving me access to the Faculty's computational systems. Thanks to Tejas Deshmukh for discussing the intricacies of this project. Additionally, the whole open-source community of Keras (Chollet 2015) deserves a big round of applause for creating a research friendly source of Deep Learning. The Kaggle community and its participants, who are all ML research enthusiasts, deserves appreciation as well. A part of the code which helped me extract data from netCDF4 files was taken from one of the forum discussions.

References

ALER, R., R. MARTÍN, J. M. VALLS, AND I. M. GALVÁN (2015): "A Study of Machine Learning Techniques for Daily Solar Energy Forecasting Using Numerical Weather Models Ricardo," *Concurrency Computation*, 29, 269–278.

BACHER, P., H. MADSEN, AND H. A. NIELSEN (2009): "Online short-term solar power forecasting," *Solar Energy*, 83, 1772–1783.

BENGIO, Y., P. SIMARD, AND P. FRASCONI (1994): "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, 5, 157–166.

BEYER, H. G., C. COSTANZO, AND D. HEINEMANN (1996): "Modifications of the heliosat procedure for irradiance estimates from satellite images," *Solar Energy*, 56, 207–212.

BEYER, H. G., C. COSTANZO, D. HEINEMANN, AND C. REISE (1994): "Short range forecast of PV energy production using satellite image analysis," in *Proc. 12th European Photovoltaic Solar Energy Conference*, Amsterdam, vol. 11, 15.

BRANCUCCI MARTINEZ-ANIDO, C., B. BOTOR, A. R. FLORITA, C. DRAXL, S. LU, H. F. HAMANN, AND B. M. HODGE (2016): "The value of day-ahead solar power forecasting improvement," *Solar Energy*, 129, 192–203.

CANO, D., J. MONGET, M. ALBUISSON, H. GUILLARD, N. REGAS, AND L. WALD (1986): "A method for the determination of the global solar radiation from meteorological satellite data," *Solar Energy*, 37, 31–39.

CAO, J. AND X. LIN (2008): "Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks," *Energy Conversion and Management*, 49, 1396–1406.

CHEN, C., S. DUAN, T. CAI, AND B. LIU (2011): "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Solar Energy*, 85, 2856–2870.

CHEN, J. AND N. S. CHAUDHARI (2005): "Protein secondary structure prediction with bidirectional lstm networks," in *International Joint Conference on Neural Networks: Post-Conference Workshop on Computational Intelligence Approaches for the Analysis of Bio-data (CI-BIO)(August 2005)*.

CHOLLET, F. (2015): "keras," \url{https://github.com/fchollet/keras}.

CHOWDHURY, B. H. (1990): "Short-Term Prediction of Solar Irradiance Using Time-Series Analysis," *Energy Sources*, 12, 199–219.

COCOCCIONI, M., E. D'ANDREA, AND B. LAZZERINI (2011): "24-Hour-ahead forecasting of

energy production in solar PV systems," *International Conference on Intelligent Systems Design and Applications, ISDA*, 1276–1281.

CORNARO, C., M. PIERRO, AND F. BUCCI (2015): "Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast," *Solar Energy*, 111, 297–312.

DIAGNE, M., M. DAVID, P. LAURET, J. BOLAND, AND N. SCHMUTZ (2013): "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, 27, 65–76.

DING, M., L. WANG, AND R. BI (2011): "An ANN-based approach for forecasting the power output of photovoltaic system," *Procedia Environmental Sciences*, 11, 1308–1315.

DUDHIA, J., D. GILL, K. MANNING, W. WANG, C. BRUYERE, S. KELLY, AND K. LACKEY (2005): "PSU/NCAR Mesoscale modeling system tutorial class notes and user's guide: MM5 modeling system version 3," *National Center for Atmospheric Research*.

DUFFY, A., M. ROGERS, AND L. AYOMPE (2015): *Renewable energy and energy efficiency: assessment of projects and policies*, John Wiley & Sons.

ECK, D. AND J. SCHMIDHUBER (2002): "Finding temporal structure in music: Blues improvisation with LSTM recurrent networks," in *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*, IEEE, 747–756.

EKICI, B. B. (2014): "A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems," *Measurement*, 50, 255–262.

ELIZONDO, D., G. HOOGENBOOM, AND R. MCCLENDON (1994): "Development of a Neural Network Model to Predict Daily Solar Radiation," *Agricultural and Forest Meteorology*, 71, 115–132.

ESPINAR, B., J. L. AZNARTE, R. GIRARD, A. M. MOUSSA, AND G. KARINIOTAKIS (2010): "Photovoltaic Forecasting: A state of the art," *5th European PV-Hybrid and Mini-Gird Conference*, 33, 250–255.

EUGENIA KALNAY (2003): *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press.

FERNÁNDEZ, Á., Y. GALA, AND R. DORRONSORO (2014): "Machine Learning Prediction of Large Area Photovoltaic Energy Production," *Proc. of the Joint ECML/PKDD 2014 Workshops*.

FONSECA J., J. G., T. OOZEKI, H. OHTAKE, K. ICHI SHIMOSE, T. TAKASHIMA, AND K. OGIMOTO (2014): "Regional forecasts and smoothing effect of photovoltaic power gener-

ation in Japan: An approach with principal component analysis," *Renewable Energy*, 68, 403–413.

GALA, Y., Á. FERNÁNDEZ, J. DÍAZ, AND J. R. DORRONSORO (2016): "Hybrid machine learning forecasting of solar radiation values," *Neurocomputing*, 176, 48–59.

GERS, F. A. AND E. SCHMIDHUBER (2001): "LSTM recurrent networks learn simple context-free and context-sensitive languages," *IEEE Transactions on Neural Networks*, 12, 1333–1340.

GERS, F. A., J. SCHMIDHUBER, AND F. CUMMINS (2000): "Learning to forget: Continual prediction with LSTM," *Neural computation*, 12, 2451–2471.

GERS, F. A., N. N. SCHRAUDOLPH, AND J. SCHMIDHUBER (2002): "Learning precise timing with LSTM recurrent networks," *Journal of machine learning research*, 3, 115–143.

GLOROT, X. AND Y. BENGIO (2010): "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 9, 249–256.

GOODFELLOW, I., Y. BENGIO, AND A. COURVILLE (2016): *Deep Learning*, MIT Press.

GRAVES, A. (2012): *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer.

GRAVES, A., N. BERINGER, AND J. SCHMIDHUBER (2005): "Rapid retraining on speech data with lstm recurrent networks," *Technical Report IDSIA-09-05, IDSIA*.

HAMILL, T., G. BATES, J. WHITAKER, D. MURRAY, M. FIORINO, AND T. GALARNEAU (2013): "A Description of the 2nd Generation NOAA Global Ensemble Reforecast Data Set," *NOAA Earth System Research Lab, Physical Sciences Division Bouder, Colorado, USA*, 10.

HAMMER, A., D. HEINEMANN, C. HOYER, R. KUHLEMANN, E. LORENZ, R. MÜLLER, AND H. G. BEYER (2003): "Solar energy assessment using remote sensing technologies," *Remote Sensing of Environment*, 86, 423–432.

HAMMER, A., D. HEINEMANN, E. LORENZ, AND B. LÜCKEHE (1999): "Short-term forecasting of solar radiation: a statistical approach using satellite data," *Solar Energy*, 67, 139–150.

HAND, D. J. (2006): "Classifier Technology and the Illusion of Progress," *Statistical Science*, 21, 1–14.

HECHT-NIELSEN, R. (1989): "Theory of the backpropagation neural network," in *Neural Networks, 1989. IJCNN., International Joint Conference on*, IEEE, 593–605.

HEINEMANN, D., E. LORENZ, AND M. GIRODO (2006): "Forecasting of solar radiation,"

Solar Energy Resource Management for Elitricity Generation from Local Level to Global Scale, 83–94.

HOCHREITER, S. (1991): "Untersuchungen zu dynamischen neuronalen Netzen," *Master's thesis, Institut fur Informatik, Technische Universitat, Munchen*.

HOCHREITER, S., Y. BENGIO, P. FRASCONI, AND J. SCHMIDHUBER (2001): "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," .

HOCHREITER, S., M. HEUSEL, AND K. OBERMAYER (2007): "Fast model-based protein homology detection without alignment," *Bioinformatics*, 23, 1728–1736.

HOCHREITER, S. AND J. SCHMIDHUBER (1997): "Long short-term memory," *Neural computation*, 9, 1735–1780.

HOFFMANN, W. (2011): "Solar photovoltaics competing in the energy sector," *8th PVSEC*, 1–38.

HORNIK, K., M. STINCHCOMBE, AND H. WHITE (1989): "Multilayer feedforward networks are universal approximators," *Neural networks*, 2, 359–366.

INMAN, R. H., H. T. C. PEDRO, AND C. F. M. COIMBRA (2013): "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, 39, 535–576.

INTERNATIONAL ENERGY AGENCY (2016): "Next Generation Wind and Solar Power," *International Energy Agency books online*.

JAMES, G., D. WITTEN, T. HASTIE, AND R. TIBSHIRANI (2000): *An introduction to Statistical Learning*, vol. 7, Springer Texts.

KEMMOKU, Y., S. ORITA, S. NAKAGAWA, AND T. SAKAKIBARA (1999): "Daily insolation forecasting using a multi-stage neural network," *Solar Energy*, 66, 193–199.

KINGMA, D. P. AND J. L. BA (2015): "Adam: a Method for Stochastic Optimization," *International Conference on Learning Representations 2015*, 1–15.

KOSTYLEV, V. AND A. PAVLOVSKI (2011): "Solar Power Forecasting Performance - Towards Industry Standards," *1st Int. Workshop on the Integration of Solar Power into Power Systems*.

KÜHNERT, J., E. LORENZ, AND D. HEINEMANN (2013): *Satellite-Based Irradiance and Power Forecasting for the German Energy Market*, Elsevier Inc.

LARSON, V. E. (2013): "Forecasting Solar Irradiance with Numerical Weather Prediction Models," *Solar Energy Forecasting and Resource Assessment*, 299–318.

LE CUN, Y., D. TOURESKY, G. HINTON, AND T. SEJNOWSKI (1988): *A theoretical framework for back-propagation*, Morgan Kaufmann, 21–28.

LEVA, S., A. DOLARA, F. GRIMACCIA, M. MUSSETTA, AND E. OGLIARI (2017): "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power," *Mathematics and Computers in Simulation*, 131, 88–100.

LINARES-RODRÍGUEZ, A., J. A. RUIZ-ARIAS, D. POZO-VÁZQUEZ, AND J. TOVAR-PESCADOR (2011): "Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks," *Energy*, 36, 5356–5365.

LIWICKI, M., A. GRAVES, H. BUNKE, AND J. SCHMIDHUBER (2007): "A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks," in *Proc. 9th Int. Conf. on Document Analysis and Recognition*, vol. 1, 367–371.

LONG, H., Z. ZHANG, AND Y. SU (2014): "Analysis of daily solar power prediction with data-driven approaches," *Applied Energy*, 126, 29–37.

LORENZ, E., A. HAMMER, AND D. HEINEMANN (2004): "Short term forecasting of solar radiation based on satellite data," *Proc. ISES Europe Solar Congress*.

LORENZ, E. AND D. HEINEMANN (2012): *Prediction of solar irradiance and photovoltaic power*, vol. 1, Elsevier Ltd.

LORENZ, E., J. HURKA, D. HEINEMANN, AND H. G. BEYER (2009a): "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2, 2–10.

LORENZ, E., J. REMUND, S. C. MÜLLER, W. TRAUNMÜLLER, G. STEINMAURER, D. POZO, J. ANTONIO, V. L. FANEGO, L. RAMIREZ, M. G. ROMEO, C. KURZ, L. M. POMARES, AND C. G. GUERRERO (2009b): "Benchmarking of diiňĂerent approaches to forecast solar irradiance," in *Proceedings of the 24th European Photovoltaic Solar Energy Conference*, 21–25.

LORENZ, E., T. SCHEIDSTEGER, J. HURJA, D. HEINEMANN, AND C. KURZ (2010): "Regional PV power prediction for improved grid integration," *Prog. Photovolt: Res. Appl.*, 19, 757–771.

LYNCH, P. (2008): "The origins of computer weather prediction and climate modeling," *Journal of Computational Physics*, 227, 3431–3444.

MAAS, A. L., A. Y. HANNUN, AND A. Y. NG (2013): "Rectifier Nonlinearities Improve Neural Network Acoustic Models," *Proceedings of the 30 th International Conference on Machine Learning*, 28, 6.

MARQUEZ, R. AND C. F. M. COIMBRA (2011): "Forecasting of global and direct solar irradi-

ance using stochastic learning methods, ground experiments and the NWS database," *Solar Energy*, 85, 746–756.

MCGOVERN, A., D. J. GAGNE, J. BASARA, T. M. HAMILL, AND D. MARGOLIN (2015): "Solar energy prediction : An international contest to initiate interdisciplinary research on compelling meteorological problems," *Bulletin of the American Meteorological Society*, 96, 1388–1393.

MELLIT, A. AND S. A. KALOGIROU (2008): "Artificial intelligence techniques for photovoltaic applications: A review," *Progress in Energy and Combustion Science*, 34, 574–632.

MELLIT, A., S. A. KALOGIROU, L. HONTORIA, AND S. SHAARI (2009): "Artificial intelligence techniques for sizing photovoltaic systems: A review," *Renewable and Sustainable Energy Reviews*, 13, 406–419.

MOHANDES, M., S. REHMAN, AND T. HALAWANI (1998): "Estimation of global solar radiation using artificial neural networks," *Renewable Energy*, 14, 179–184.

MORENO, A., M. A. GILABERT, AND B. MARTÍNEZ (2011): "Mapping daily global solar irradiation over Spain: A comparative study of selected approaches," *Solar Energy*, 85, 2072–2084.

MÜLLER, S. C. AND J. REMUND (2010): "Advances in radiation forecast based on regional weather models MM5 and WRF," in *Proceedings of the 25th EUPVSEC Conference*, 6–9.

NAIR, V. AND G. E. HINTON (2010): "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proceedings of the 27th International Conference on Machine Learning*, 807–814.

PAOLI, C., C. VOYANT, M. MUSELLI, AND M. L. NIVET (2010): "Forecasting of preprocessed daily solar radiation time series using neural networks," *Solar Energy*, 84, 2146–2160.

PEEL, M. C., B. L. FINLAYSON, AND T. A. MCMAHON (2007): "Updated world map of the K{ö}ppen-Geiger climate classification," *Hydrology and earth system sciences discussions*, 4, 439–473.

PEREZ, R., S. KIVALOV, J. SCHLEMMER, K. HEMKER, D. RENNÉ, AND T. E. HOFF (2010): "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy*, 84, 2161–2172.

PEREZ, R., E. LORENZ, S. PELLAND, M. BEAUHARNOIS, G. VAN KNOWE, K. HEMKER, D. HEINEMANN, J. REMUND, S. C. MÜLLER, W. TRAUNMÜLLER, G. STEINMAUER, D. POZO, J. A. RUIZ-ARIAS, V. LARA-FANEGO, L. RAMIREZ-SANTIGOSA, M. GASTON-ROMERO, AND

L. M. POMARES (2013): "Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe," *Solar Energy*, 94, 305–326.

PEREZ, R., K. ZWEIBEL, AND T. E. HOFF (2011): "Solar power generation in the US: Too expensive, or a bargain?" *Energy Policy*, 39, 7290–7297.

PODESTÁ, G. P., L. NÚÑEZ, C. A. VILLANUEVA, AND M. A. SKANSI (2004): "Estimating daily solar radiation in the Argentine Pampas," *Agricultural and Forest Meteorology*, 123, 41–53.

RAZA, M. Q., M. NADARAJAH, AND C. EKANAYAKE (2016): "On recent advances in PV output power forecast," *Solar Energy*, 136, 125–144.

REIKARD, G. (2009): "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, 83, 342–349.

RUMELHART, D. E., G. E. HINTON, AND R. J. WILLIAMS (1986): "Learning representations by back-propagating errors," *Nature*, 323, 533–536.

SALCEDO-SANZ, S., C. CASANOVA-MATEO, A. PASTOR-SÁNCHEZ, AND M. SÁNCHEZ-GIRÓN (2014): "Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization âĂŞ Extreme Learning Machine approach," *Solar Energy*, 105, 91–98.

SCHMIDHUBER, J. (2015): "Deep learning in neural networks: An overview," *Neural Networks*, 61, 85–117.

SCHROEDTER-HOMSCHEIDT, M. (2016): "The Copernicus Atmosphere Monitoring Service (CAMS) Radiation Service in a nutshell," Tech. rep.

SENGUPTA, M., A. HABTE, S. KURTZ, A. DOBOS, S. WILBERT, E. LORENZ, T. STOFFEL, D. RENNÉ, C. GUEYMARD, D. MYERS, S. WILCOX, P. BLANC, AND R. PEREZ (2015): "Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications," *Technical Report, NREL/TP-5D00-63112*, 1–255.

SENGUPTA, M., A. WEEKLEY, A. HABTE, A. LOPEZ, AND C. MOLLING (2012): "Validation of the National Solar Radiation Database (NSRDB) Preprint," *Nrel*.

SIEGELMANN, H. T. AND E. D. SONTAG (1992): "On the computational power of neural nets," in *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, 440–449.

SKAMAROCK, W. C., J. B. KLEMP, J. DUDHIA, D. O. GILL, D. M. BARKER, W. WANG, AND J. G. POWERS (2005): "A description of the advanced research WRF version 2," Tech. rep., DTIC Document.

SRIVASTAVA, N., G. HINTON, A. KRIZHEVSKY, I. SUTSKEVER, AND R. SALAKHUTDINOV (2014): "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, 15, 1929–1958.

TADESSE, A., A. KANKIEWICZ, R. PEREZ, AND P. LAURET (2016): "Day Ahead Irradiance Forecast Variability Characterization Using Satellite Data," *43rd IEEE Photovoltaic Specialists Conference*.

VOYANT, C., G. NOTTON, S. KALOGIROU, M.-L. NIVET, C. PAOLI, F. MOTTE, AND A. FOUILLOY (2017): "Machine Learning methods for solar radiation forecasting: a review," *Renewable Energy*, 105, 569–582.

VOYANT, C., P. RANDIMBIVOLOLONA, M. L. NIVET, C. PAOLI, AND M. MUSELLI (2014): "Twenty four hours ahead global irradiation forecasting using multi-layer perceptron," *Meteorological Applications*, 21, 644–655.

WAINER, H. (1976): "Estimating Coefficients in Linear Models : It Don't Make No Nevermind," *Psychological Bulletin*, 83, 213–217.

WANG, G., Y. SU, AND L. SHU (2016): "One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models," *Renewable Energy*, 96, 469–478.

WERBOS, P. J. (1981): "Applications of Advances in Nonlinear Sensitivity Analysis," in *Proceedings of the 10th IFIP Conference, 31.8 - 4.9, NYC*, 762–770.

WIRTH, H. (2017): "Recent facts about photovoltaics in Germany," *Fraunhofer ISE*, 92.

YONA, A., T. SENJYU, A. Y. SABER, T. FUNABASHI, H. SEKINE, AND C. H. KIM (2007): "Application of neural network to one-day-ahead 24 hours generating power forecasting for photovoltaic system," *2007 International Conference on Intelligent Systems Applications to Power Systems, ISAP*, 442–447.

ZHAOXUAN, L., R. SM MAHBOBUR, R. VEGA, AND B. DONG (2016): "A Hierarchical Approach Using Machine Learning Methods in Solar Photovoltaic Energy Production Forecasting," *Energies*, 9, 55.

A Figures

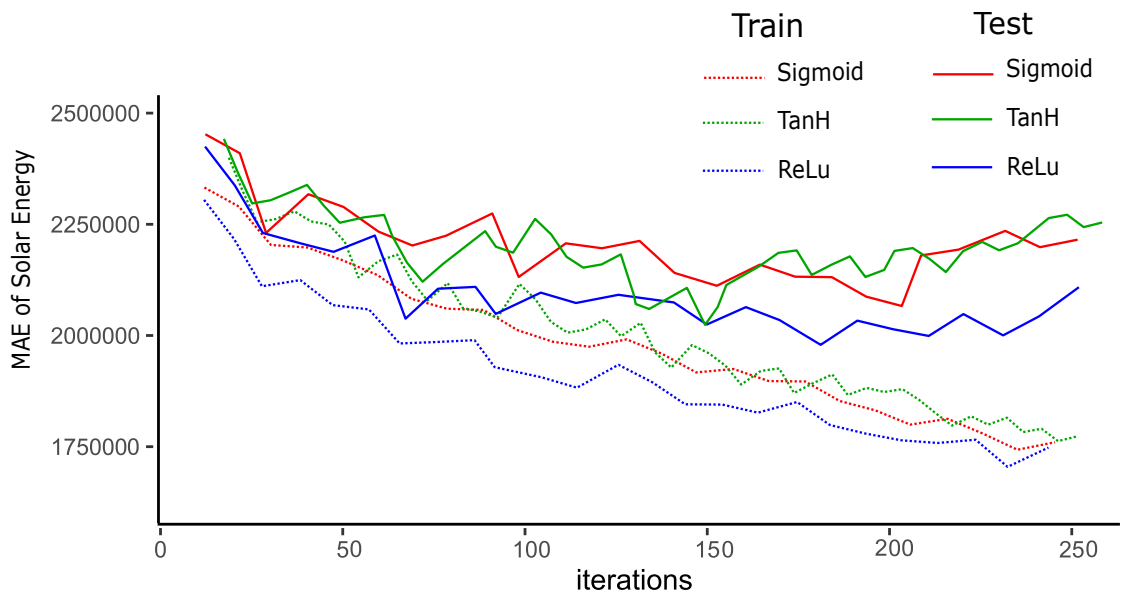


Figure 14: A schematic view of performances of LSTM using different activation functions

B Tables

Table 7: Overall Improvement in RMSE and MAE using LSTM over GBR and FFNN

	RMSE		MAE	
Location	**FFNN**	**GBR**	**FFNN**	**GBR**
Europe				
Hohenpeissenberg	2.5%	1.8%	7.4%	5.8%
Cabauw	9.4%	9.4%	6.7%	3.0%
De Bilt	4.8%	0.2%	2.2%	-1.5%
Grossenzersdorf	-0.8%	11.8%	3.3%	11.9%
Sonnblick	-1.9%	-6.0%	2.0%	-1.6%
Wien	11.5%	18.0%	13.3%	17.7%
Carpentras	11.0%	7.0%	8.2%	7.5%
Palaiseau	-7.1%	-18.6%	-0.5%	-4.2%
Santander	3.5%	5.0%	4.1%	5.1%
La Coruna	7.3%	0.9%	12.2%	4.5%
Alicante	0.7%	1.8%	3.6%	13.2%
Madrid	7.2%	-3.4%	18.0%	3.4%
Basel	12.6%	9.5%	12.2%	9.9%
Zurich	8.6%	1.7%	12.5%	2.8%
Geneva	0.4%	0.2%	5.9%	3.9%
Davos	-1.8%	-1.5%	1.5%	2.7%
US				
Desert Rock	12.3%	5.6%	12.3%	5.6%
Bondville	13.8%	4.8%	13.8%	4.8%
Fort Peck	9.5%	6.5%	9.5%	6.5%
Goodwin Creek	15.3%	2.8%	15.3%	2.8%
Penn State	9.5%	3.8%	9.5%	3.8%

Table 8: Month wise improvement in RMSE score using LSTM over GBR and FFNN

	Jan		Feb		Mar		Apr		May		Jun		Jul		Aug		Sep		Oct		Nov		Dec	
Location	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**	**FFNN**	**GBR**
Europe																								
Hohenpeissenberg	20%	22%	18%	13%	23%	11%	8%	9%	2%	-10%	14%	3%	-17%	1%	7%	7%	2%	1%	-24%	-2%	7%	10%	20%	23%
Cabauw	15%	8%	13%	-3%	-13%	-8%	1%	11%	21%	8%	22%	19%	18%	5%	11%	21%	5%	29%	-33%	-45%	5%	-17%	13%	11%
De Bilt	8%	-12%	-2%	-12%	-8%	-16%	-5%	7%	11%	-3%	14%	5%	27%	3%	9%	14%	16%	16%	-92%	-35%	0%	-1%	-12%	1%
Grossenzersdorf	-7%	-18%	-11%	-2%	-19%	2%	10%	18%	-1%	6%	20%	16%	23%	36%	-6%	-18%	0%	19%	-25%	9%	8%	27%	23%	10%
Sonnblick	5%	-4%	-5%	-22%	-13%	11%	12%	2%	1%	-5%	13%	4%	8%	7%	8%	2%	-7%	5%	-77%	-123%	16%	13%	-2%	9%
Wien	14%	-1%	2%	8%	-34%	-5%	3%	21%	28%	24%	41%	26%	41%	40%	-10%	16%	13%	22%	-36%	1%	17%	32%	54%	31%
Carpentras	12%	-9%	11%	17%	19%	20%	9%	11%	9%	-2%	13%	-2%	-1%	16%	-14%	6%	-7%	-9%	28%	-12%	14%	29%	5%	12%
Palaiseau	5%	12%	16%	28%	-8%	-10%	10%	-2%	14%	-7%	6%	-9%	7%	12%	0%	-16%	17%	9%	-111%	-142%	1%	9%	-1%	-12%
Santander	4%	8%	-13%	0%	-13%	2%	-2%	-3%	22%	5%	-2%	10%	23%	24%	0%	2%	7%	16%	-21%	-50%	1%	0%	-11%	-5%
La Coruna	14%	25%	34%	13%	16%	-10%	-4%	3%	13%	0%	19%	6%	-11%	-2%	-5%	2%	8%	-2%	7%	-8%	8%	9%	-20%	-1%
Alicante	-5%	-7%	8%	-5%	2%	22%	1%	6%	-4%	-9%	0%	-12%	17%	13%	20%	28%	19%	21%	-15%	-29%	-5%	15%	-30%	12%
Madrid	-1%	9%	-6%	-4%	16%	20%	9%	21%	26%	3%	33%	7%	11%	19%	5%	-44%	11%	-13%	-28%	-89%	4%	-7%	-6%	11%
Basel	-1%	9%	-2%	1%	22%	-6%	16%	-4%	7%	19%	16%	3%	21%	12%	7%	21%	15%	-2%	13%	1%	-15%	16%	17%	21%
Zurich	15%	-5%	11%	-21%	19%	5%	-1%	-11%	-5%	-1%	12%	6%	4%	9%	10%	19%	15%	-13%	32%	8%	13%	-13%	-5%	-19%
Geneva	8%	-12%	0%	-18%	-2%	19%	-15%	8%	17%	1%	-9%	-4%	11%	4%	4%	4%	13%	19%	4%	15%	7%	-13%	-10%	-7%
Davos	3%	9%	7%	0%	24%	6%	3%	12%	2%	-1%	5%	-11%	-18%	-4%	6%	14%	10%	15%	-47%	-49%	12%	10%	5%	14%
Average	**7%**	**2%**	**5%**	-1%	**2%**	**4%**	**3%**	**7%**	**10%**	**2%**	**14%**	**4%**	**10%**	**12%**	**3%**	**5%**	**9%**	**8%**	**-27%**	**-34%**	**6%**	**7%**	**2%**	**7%**
US																								
Desert Rock	-3%	25%	8%	3%	-4%	6%	-5%	-7%	-14%	3%	25%	26%	1%	0%	8%	19%	-35%	10%	26%	-10%	2%	-1%	-2%	1%
Bondville	3%	-2%	19%	-2%	38%	8%	6%	12%	16%	3%	33%	-5%	20%	-10%	40%	14%	42%	20%	32%	-2%	-5%	0%	5%	-2%
Fort Peck	10%	-12%	6%	-5%	2%	7%	2%	4%	10%	-4%	25%	9%	10%	19%	15%	8%	12%	20%	32%	1%	8%	17%	-5%	0%
Goodwin Creek	11%	-11%	28%	1%	-3%	0%	7%	19%	24%	-10%	-9%	-9%	-5%	-13%	11%	11%	-2%	10%	5%	14%	21%	18%	-4%	12%
Penn State	35%	-9%	-4%	-21%	14%	0%	-9%	7%	-15%	8%	21%	11%	33%	15%	16%	13%	8%	-1%	21%	-19%	1%	3%	23%	-1%
Average	**11%**	-2%	**11%**	-5%	**9%**	**4%**	**0%**	**7%**	**4%**	**0%**	**19%**	**6%**	**12%**	**2%**	**18%**	**13%**	**5%**	**12%**	**23%**	-3%	**6%**	**7%**	**3%**	**2%**

Table 9: Abbreviations of models from Table 1

Abbr	**Model Name**
GAM	Generalized Additive Model
RBF	Radial Bias Function
DRWNN	Diagonal Recurrent Wavelet Network
RLS	Recursive Least Square
AP	ångström-Prescott equation
K-RR	Kernel Ridge Regression
NP	Naïve Predictor
MC	Markov Chain
K-NN	K-Nearest Neighbor
AR	Auto-Regression
BC	Bristow-Campbel
ARMA	Autoregressive Moving Average
ELM	Extreme Learning Machine
SVR	Support Vector Regression
RFR	Random Forest Regression
PF-LR	Partial Functions- Linear Regression

Table 10: Meta-parameters used for Grid Search

Algorithm	**Meta-Parameter**	**Values**
LSTM	Neurons in 1st Layer	c(10,20,40,60,80,100,110,120,140,150)
	Neurons in 2nd Layer	c(10,20,40,60,80,100,110,120,140,150)
FFNN	Neurons in 1st Layer	c(10,20,40,60,80,100,110,120,140,150)
	Neurons in 2nd Layer	c(10,20,40,60,80,100,110,120,140,150)
GBR	Shrinkage	c(0.005,0.01,0.05,0.1)
	Minobsinnode	c(8,9,10,11,12)
	Interaction depth	c(8,9,10,11,12)

Declaration of Authorship

I hereby confirm that I have authored this Master's thesis independently and without use of others than the indicated sources. All passages which are literally or in general matter taken out of publications or other sources are marked as such.

Shikhar Srivastava

Berlin, February 24, 2017