2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Hybrid Model Analysis and Validation for PV energy production forecasting

A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, E. Ogliari
Politecnico di Milano – Department of Energy, via Lambruschini 34, Milano – Italy

Abstract— **In this paper a forecasting method for the Next Day's energy production forecast is proposed with respect to photovoltaic plants. A new hybrid method PHANN (Physical Hybrid Artificial Neural Network) based on Artificial Neural Network (ANN) and basic Physical constraints of the PV plant, is presented and compared with an ANN standard method. Furthermore, the accuracy of the two methods have been studied in order to better understand the intrinsic error committed by the PHANN, reporting some numerical results. This computing-based hybrid approach is proposed for PV energy forecasting in view of optimal usage and management of RES in future smart grid applications.**

I. INTRODUCTION

The sustainable usage of renewable sources has become a critical issue which will represent a key challenge in the next future. The increase in demand due to developing countries and the likely recovery of economy in developed countries, even in Europe, promises to stimulate new research on renewable energy resources (RES), especially oriented to advanced grids able to optimally manage the energy delivery in a distributed system. The number of plants producing electricity has been enormously increased and most of the installations are decentralized. Thus the old centralized model of electric generation tends towards a mixed system. Challenges of controlling and maintaining energy from intermittent sources involve many research topics like efficiency, reliability, safety and stability of the grid: all these aspects can take advantage of the ability to forecast energy flows [1, 2].

In this context evolutionary computation and optimization algorithms can be extremely useful in the next smart grid development. Besides, forecasting tool related to neural systems and computational intelligence can play a fundamental role both in energy production and demand side consumption. The predictive task is surely challenging but the relative technical and economic benefit cannot be neglected for dynamic optimization of grid operations and market transactions.

In recent years several short-term power forecasting models related to PV plants have been presented and they can be generally classified into physical, statistical and hybrid methods. Some of these models started to predict solar radiation [3-5], some others have been focused on hourly power generation forecasting [6-9]. The most applied techniques in these forecasting models are Artificial Neural Networks (ANNs) [10] even if some works use simple physical methods. The authors previously started to evaluate artificial intelligence methods for PV plant production for integration in smart energy systems [11] with a different perspective.

In this paper the aim is to present a novel hybrid model that combines a soft-computing model based on ANN and a physical model short-term power forecasting of a PV plant. The paper is organized as follows: in Section 2 a brief review of the hourly energy production forecasting methods is presented. In Section 3, the new proposed hybrid methods is described. In Section 4 some error indexes that can be used to evaluate the performances of the forecasting models are defined. In Section 5 the power prediction, in terms of hourly error is presented.

Finally, we conclude and outline additional research directions for next future work.

II. ENERGY FORECAST MODELS

RES energy production forecasting methods are commonly divided in different categories: Physic, Stochastic and Hybrid. An analysis of state-of-the-art approaches is proposed in [12] but it is also important to further develop methods based on weather forecast [13].

In physical models the ability of a RES plant to convert the introduced meteorological resources into electrical power are summarized by a physical-analytical model. Statistical methods are based on the concept of persistence, or stochastic time series. Nowadays the most common approach to forecast a time series' future values approach is the use of machine learning methods. These methods learn to recognize patterns in data using training data sets. Any combination of two or more of the previously described methods can be defined as an hybrid model.

III. THE NEW PROPOSED HYBRID METHOD

In order to make the hourly prediction of the production of a RES plant more accurate, a new Hybrid forecasting system based on ANN that incorporates some of the physical model constraints (see Fig. 1) was developed and it was called PHANN (Physical Hybridized Artificial Neural Network). In particular it includes a theoretical model of the solar radiation mathematically computed according to the geographical coordinates of the PV plant site (Clear Sky Solar

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

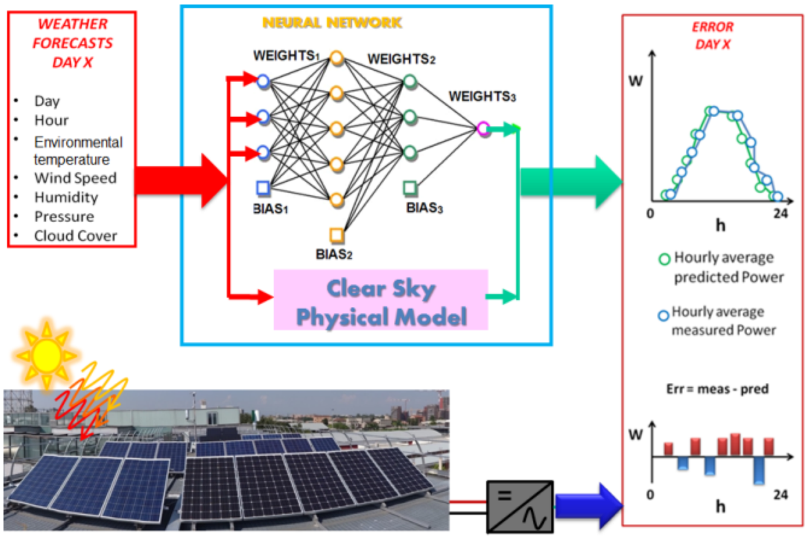

Fig. 1. Illustrative diagram of the new method applied.

Radiation Model) [14].

The block diagram of Fig. 2 illustrates the different phases carried out for the proposed procedure. It shows the main blocks (input, process and analysis) [15,16] and the fine-tuning activity performed retrospectively in order to refine and further to improve the quality of the prediction.

For the training phase, both the input and the output historical data of the PV plant are required in order to lead a supervised learning of the PHANN. Once the PHANN is trained and tuned, it can be used to provide predictions of the PV system output power by supplying only the weather forecasts as input. In this context also partial shading effects can be taken into account [17]. After this phase, the final accuracy assessment of the results should be carried out.

A. Pre-Processing and raw data Validation

About the training of the neural network and the analysis of the forecast errors, historical measured data acquired must always be validated. In fact the use of not reliable data generally imply an increase of the error in the prediction. Data validation is performed by evaluating the reliability of each fifteen minutes average measured and data provided by the Meteorological Service on the basis of comparisons between solar radiation, power output and theoretical radiation predicted by a mathematical model, the "Clear Sky" solar radiation [14].

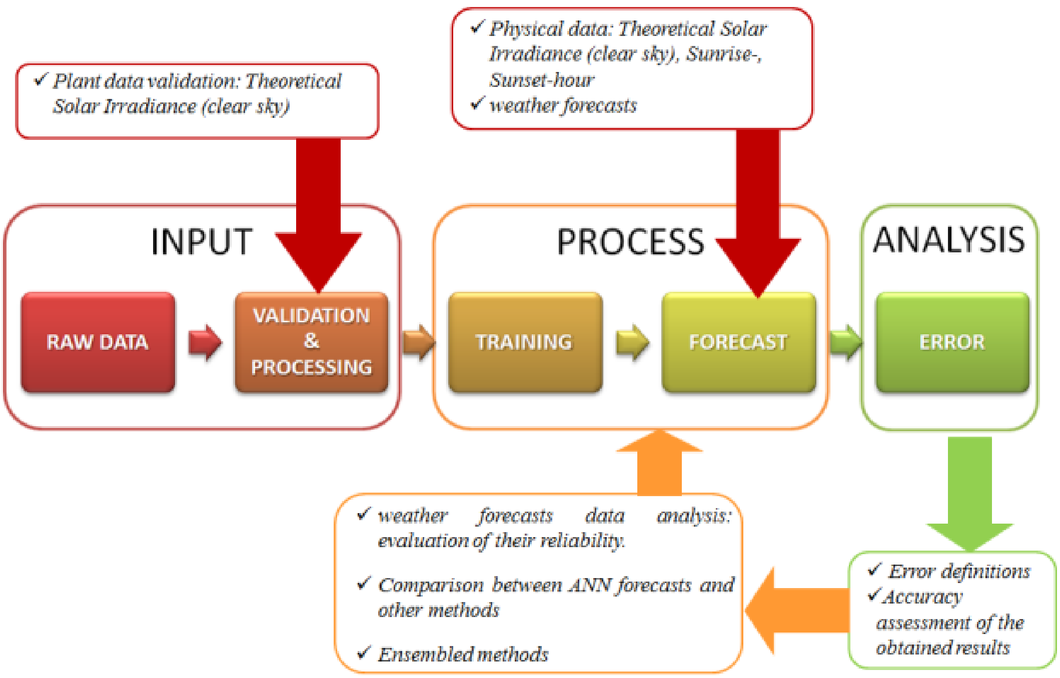

Fig. 2. Block diagram of the hybrid method used for forecasting.

B. The forecasting process

Among the several existing methods, the proposed procedure is based on a physical approach combined with an ANN trained with classical algorithms. The considered neural network has a classical structure (MLP) with an Error Back Propagation training procedure.

The input of the tool are the weather forecasts provided by the meteorological service, the geographical coordinates of the site and the date and time to determine the correct sun position. The output of the tool is the predicted value of the hourly power produced by the PV plant for a given time.

Considering a generic network with n input and p output t_j and o_j respectively are the expected output and the actual output of the j neuron. The E_k error committed by the ANN on the k-th sample is defined by:

$$E_k = \frac{1}{2}\sum_{j=1}^{p}\left(t_j - o_j\right)^2 \quad (1)$$

where ½ factor has been introduced to simplify the notation. Therefore the global error E committed by the Neural Network on the whole training with m samples is:

$$E = \sum_{k=1}^{m} E_k \quad (2)$$

In the case of linear neurons E is a paraboloid in the weights space, as it is shown in Fig. 3.

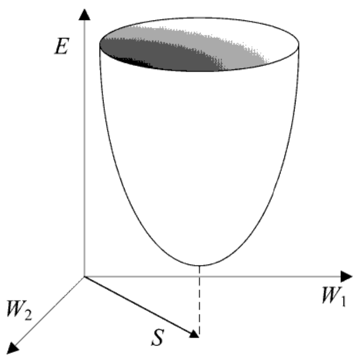

Fig. 3. Error surface E in the case of linear neurons, S is the weight vector that generates the minimum error.

In our case study neurons are sigmoidal, and the training error convergence to the minimum is shown for example in Fig. 4.

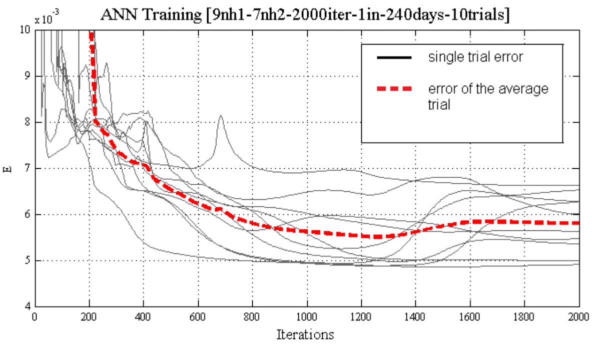

Fig. 4. ANN training set with different trials error and the error of the average trial for one of the case study.

C. Error Assessment

According to the error definitions of the Output Power Forecast in comparison with the Measured one exposed in the next paragraph, an error Assessment is performed to identify the error with the lowest order of magnitude.

D. Fine tuning

Fine tuning step introduces, in addition to the clear sky model, some additional Physical Constraints which are useful to limit the effect of the forecasting error in those hour of low solar radiation. These Physical Constraints are imposed to some specific hourly average parameters: G_f, G_t, G_m, P_m, $Affi_h$, where G_f is the global tilted forecast irradiance (W/m²) and G_t is theoretical tilted irradiance (W/m²), G_m is the measured irradiance on the PV panels (W/m²), P_m is the measured AC power from the plant (kW) and $Affi_h$ is the availability in the hour of the every 15 minutes sampled data. It may have values included between "0" and "1".

IV. Error Definitions

In order to correctly define the accuracy of the prediction and the relative error, it is necessary to define the error indexes that can be used to evaluate the performances of the forecasting models. Some of this definitions come from statistics, others are introduced by regulatory authority for market issues.

The starting point reference is the hourly error e_h defined as the difference between the average power produced (measured) in the "h" hour $P_{m,h}$ and the given prediction $P_{p,h}$ provided by the forecasting model:

$$e_h = P_{m,h} - P_{p,h} \text{ [Wh]} \quad (3)$$

From this basic definition, other definitions can be introduced.

Absolute hourly error $e_{h,abs}$ which is the absolute value of the previous definition (e_h can give both positive and negative values):

$$e_{h,abs} = |e_h| \text{ [Wh]} \quad (4)$$

Time error percentage could be $e_{\%,p}$, if it is based on the hourly output expected power hour $P_{p,h}$ or could be $e_{\%,m}$, if it is based on the hourly output measured power hour $P_{m,h}$:

$$e_{\%,p} = \frac{|e_h|}{P_{p,h}} \cdot 100[\%] \quad (5)$$

Normalized mean absolute error $NMAE\%$, based on net capacity of the plant (C):

$$NMAE_{\%} = \frac{1}{N} \cdot \sum_{h=1}^{N} \frac{|P_{m,h} - P_{p,h}|}{C} \cdot 100 \quad (6)$$

where C is the "net capacity of the plant". For this indicator the rated power of the PV system was considered as C.

Weighted mean absolute error $WMAE\%$, based on power:

$$WMAE_{\%} = \frac{\sum_{h=1}^{N} |P_{\mathrm{m,h}} - P_{\mathrm{p,h}}|}{\sum_{h=1}^{N} P_{\mathrm{m,h}}} \cdot 100 \quad (7)$$

Normalized root mean square error *nRMSE*, based on the maximum observed power output $P_{\mathrm{m,h}}$ [A]:

$$nRMSE = \frac{\sqrt{\frac{\sum_{h=1}^{N} |P_{\mathrm{m,h}} - P_{\mathrm{p,h}}|^2}{N}}}{\max(P_{\mathrm{m,h}})} \cdot 100 \quad (8)$$

V. Case Study

This section describes the forecast achieved by comparing the results of the two models:

i. ANN: Artificial Neural Network without Clear Sky and Fine Tuning (Physical Constraints)
ii. PHANN: Physical Hybridized Artificial Neural Network.

The performed simulations had the main objective to compare the accuracy of the forecasts first using the ANN method alone and then combined with the physical constraints in the PHANN. Moreover the error of the two methods has been estimated by replacing the forecast values with the actual solar radiation samples measured by the monitoring system. Finally the results of some significant days are presented. The procedure has been then developed varying both iterations and training period to improve the effectiveness of the proposed technique.

A. Characteristics of data and models

While the PV plant hourly electric power generation data are recorded with measurement equipment placed in its location, the Meteo variables data are obtained by a weather forecasting service with 72hours in advance.

With reference to the prediction of hourly production relative to twenty-four hours the analysis is performed using an ANN with 10 months dataset, 9 neurons in the first hidden layer and 7 neurons in the second hidden layer, which is the configuration that proved to be a good compromise in terms of effectiveness and time-efficiency. For each forecasting simulation, 10 trials have been performed.

B. Evaluation criteria and tests

To evaluate the performance of the forecasting ANN and PHANN models the NMAE, WMAE and nRMSE errors are calculated. Table I provides just some of the most important ones. The results refer to different ANN training settings, namely:

CASE A. ANN without Physical Constraints.
CASE B. PHANN with these Physical Constraints:

a. $G_{\mathrm{f}} > 50\mathrm{W/m^2}$ and $P_{\mathrm{m}} > 0$ kW with this forecast irradiance condition it is assumed to have a positive PV plant production, in order to make the forecast for this hour.
b. $0 > G_{\mathrm{t}} > 1\mathrm{W/m^2}$ those samples with positive nearly zero theoretical irradiance condition are rounded to $1\mathrm{W/m^2}$.
c. $Affi_{hour} \geq 0.75$ only those samples with the hourly average reliability more than 0.75 are considered.

CASE C. ANN In this case, the input dataset has $G_{\mathrm{f}} = G_{\mathrm{m}}$ as if the solar radiation forecast is replaced by the actual solar radiation measured by the monitoring system.
CASE D. PHANN with the same Physical Constraints of case A. In this one, the input dataset has $G_{\mathrm{f}} = G_{\mathrm{m}}$ as if the solar radiation forecast is replaced by the actual solar radiation measured by the monitoring system.

The aim of the last two cases (C and D) is to assess the robustness of the methods according to the weather forecast accuracy given by the Meteo service.

VI. Results

They main results of the analysis are summarized in the Table I, comparing the efficiency of the PHANN method and the ANN alone.

In particular table shows in the first two columns how all values of the errors, calculated according to the definitions applied to the whole period with 150 days of training, are lowered using the Hybrid method. The most affected one is WMAE that is reduced by 42%, while, in this forecasting scenario, NMAE and nRMSE show smaller decreases.

TABLE I
Error Definitions Applied To The Whole Period Forecast With Different ANN Settings

Tr. days: 150	2000 Iterations		5000 iterations	
CASE	A	B	C	D
NMAE%	10.7	8.95	6.52	5.51
WMAE%	46.8	27	24.1	19
nRMSE%	22.70	18.13	13.51	11.66

Table II shows the relative errors, and in this case they are lowered again using the Hybrid method and the most affected one is WMAE while NMAE and nRMSE show smaller decreases.

TABLE II
Error Definitions Applied To The Whole Period Forecast With Different ANN Settings

Tr. days: 240	2000 iterations		5000 iterations	
CASE	A	B	C	D
NMAE%	8.07	6.11	7.23	4.24
WMAE%	32.7	25.9	22.3	15.3
nRMSE%	14.52	10.51	14.59	10.54

According to reported results the hybrid forecast based on the most accurate solar radiation forecasts did not show any specific improvement.

Besides some significant days (namely Sunny Day, Partially Cloudy Day and Cloudy Day) have been preliminarily taken into account in order to evaluate the Hybrid method forecast accuracy applied to a reduced number of hourly samples. Fig. 5 shows the accuracy of the hybrid forecast, with the same settings listed before, applied to a sample day with sunny weather condition.

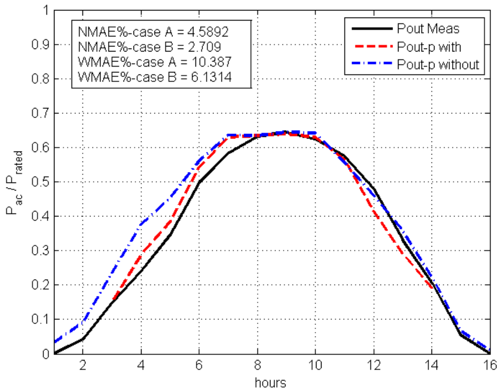

Fig. 5. Error definitions applied to the forecasts in a sunny day.

The trends of the PV plant Power output in each of the considered days shows the same behavior of the Hybrid forecasting effectiveness highlighted in the whole presented period with some different features gathered in Table III.

These figures show the remarkable importance the weather forecasts reliability plays on the energy production forecasting activity. Besides it can be noticed how the new hybrid method proposed here is more effective than traditional ones. With a focus on the first and the fourth days proposed in Table III, it can be seen that both NMAE and WMAE are significantly reduced by using this new hybrid method.

TABLE III
ERROR DEFINITIONS APPLIED TO THE FORECAST OF FOUR SIGNIFICANT DAYS WITH DIFFERENT ANN SETTINGS AND WEATHER CONDITIONS

SAMPLE	FORECASTING METHOD	NMAE%	WMAE%
Sunny Day	PHANN	2.70	6.13
	ANN	4.58	10.38
Partially Cloudy Day	PHANN	5.70	14.52
	ANN	6.58	16.77
Cloudy Day 1	PHANN	23.61	108.82
	ANN	23.05	106.25
Cloudy Day 2	PHANN	11.66	52.16
	ANN	16.04	71.74

VII. CONCLUSION

In this paper a new hybrid forecasting method, by means of artificial neural network with physical constraints is presented. The results from the error assessment, according to the error definitions here explained, lead to the conclusion that PHANN method is more accurate than the ANN one even if the quality of the data in the training set is still a key parameter to be taken in account. Besides it has been emphasized that the accuracy of these methods is strictly related to the historical data preprocess phase and to the accuracy of the historical weather forecasting data used for the training phase. The errors trend clearly shows how the accuracy in the sunny days is higher with PHANN in comparison to ANN method, while in cloudy days the overall efficiency is slightly different.

Some future improvements are therefore connected to the reliability of the weather forecasting and a more in depth study is required in order to evaluate the best suitable training time span to optimize the new procedure performance.

ACKNOWLEDGMENT

This study is based on an experimental activities carried out at Solar Tech Lab, http://www.solartech.polimi.it/.

REFERENCES

[1] M. Brenna, A. Dolara, F. Foiadelli, G. C. Lazaroiu, S. Leva, "Transient Analysis of Large Scale PV Systems with Floating DC Section", Energies 2012, 5(10), 3736-3752

[2] P. Bellezza Quater, F. Grimaccia, S. Leva, M. Mussetta, and M. Aghaei, "Light Unmanned Aerial Vehicles (UAVs) for Cooperative Inspection of PV Plants", IEEE Journal of Photovoltaics, 2014

[3] A. Mellit, A. Massi Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste", Italy, Solar Energy, Volume 84, Issue 5, May 2010, Pages 807-821.

[4] F.O. Hocaoglu, O.N. Gerek, and M. Kurban, "Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks", Solar Energy, vol. 82, no. 8, pp. 714–726, 2008.

[5] F. Wang, Z. Mi, S. Su, and H. Zhao, "Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters", Energies, vol. 5, no. 5, pp. 1355–1370, 2012.

[6] H.T.C. Pedro and C.F.M. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs", Solar Energy, vol. 86, no. 7, pp. 2017–2028, 2012.

[7] E. Izgi, A. Öztopal, B. Yerli, M.K. Kaymak, and A.D. Sahin, "Short-mid-term solar power prediction by using artificial neural networks", Solar Energy, vol. 86, no. 2, pp. 725–733, 2012.

[8] J. Shi, W.J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines", IEEE T. on Industry Applications, vol. 48, no. 3, pp. 1064–1069, 2012.

[9] T. Alamsyah, K. Sopian, and A. Shahrir, “Predicting average energy conversion of photovoltaic system in Malaysia using a simplified method”, Renewable Energy, 29(3):403 - 411, 2004.

[10] D. Caputo, F. Grimaccia, M. Mussetta, R.E. Zich, “Photovoltaic plants predictive model by means of ANN trained by a hybrid evolutionary algorithm”, Proceedings of the International Joint Conference on Neural Networks (IJCNN 2010), doi:10.1109/IJCNN.2010.5596782, July 2010.

[11] Simonov, M., Mussetta, M., Grimaccia, F., Leva, S., Zich, R.E., “Artificial intelligence forecast of PV plant production for integration in smart energy systems”, International Review of Electrical Engineering, 7 (1), pp. 3454-3460, 2012.

[12] Robert Ulbricht, Ulrike Fischer, Wolfgang Lehner, and Hilko Donker, “First Steps Towards a Systematical Optimized Strategy for Solar Energy Supply Forecasting, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases” (ECMLPKDD 2013), Prague, September 23th to 27th, 2013.

[13] F. Grimaccia, M. Mussetta, R. Zich, “Neuro-fuzzy predictive model for PV energy production based on weather forecast”, IEEE International Conference on Fuzzy Systems, doi:10.1109/FUZZY.2011.6007687, pp. 2454–2457, 2011.

[14] Richard E. Bird, “A simple solar spectral model for direct-normal and diffuse horizontal irradiance”, Solar Energy, Volume 32, Issue 4, 1984, Pages 461–471.

[15] E. Ogliari, F. Grimaccia, S. Leva, and M. Mussetta. “Hybrid predictive models for accurate forecasting in PV systems.” Energies, 6(4):1918-1929, 2013.

[16] F. Grimaccia, S. Leva, M. Mussetta, E. Ogliari, “Analysis and Validation of ANN PV power output forecasting at 24 hours ahead”, Proc. ELECTRIMACS 2014, 19th-22nd May 2014, Valencia, Spain

[17] A. Dolara, G. C. Lazaroiu, S. Leva, and G. Manzolini, “Experimental investigation of partial shading scenarios on PV (photo-voltaic) modules", Energy, vol. 55, pp. 466–475, 2013.