

PREVISÃO EM CURTO PRAZO DE ÍNDICE DE CLARIDADE UTILIZANDO REDES NEURAIS E VARIÁVEIS SAZONAIS

E. Sabino^{1,2,3}, L. Petribu^{1,2}, H. Barros^{1,2}, A. Costa^{1,2}, O. C. Vilela^{1,2}

¹ Departamento de Energia Nuclear da Universidade Federal de Pernambuco (DEN-UFPE)..Av. Prof. Luiz Freire, s/n, Recife-PE, Brasil.

² Centro de Energias Renováveis da Universidade Federal de Pernambuco (CER-UFPE). Av. da Arquitetura, s/n, Recife-PE, Brasil.

³ Instituto de Educação, Ciência e Tecnologia de Pernambuco (IFPE), Campus Pesqueira, BR 232, km 208, Prado, Pesqueira-PE, Brasil.
E-mail: edson.sabino@pesqueira.ifpe.edu.br

Recibido 26/07/17, aceptado 27/09/17

RESUMEN: Este trabajo busca contribuir a la predicción de la radiación solar horaria en horizontes de hasta 12 horas adelante. Para ello, han sido empleadas Redes Neuronales (ANNs) del tipo Perceptron Multicapa. A su vez, dichas redes han empleado, como variables regresoras, datos horarios de Índice de Claridad (Kt) y variables determinísticas estacionales con el propósito de mejorar la comprensión en cuanto a las características de estacionalidad diaria y anual de la serie de Kt para un emplazamiento en Petrolina-PE-Brasil. El objetivo es comparar los resultados logrados con el empleo de dichas variables estacionales (que, entre otros aspectos, tienen características de variables difusas - fuzzy variables) con aquellos logrados por otros modelos reportados por diferentes autores como las mismas ANNs sin las variables estacionales, los Modelos Autoregresivos y la Persistencia. El error (nRMSD) del modelo final ha variado entre 17,8% y 25,7% para horizontes de predicción desde 1 hasta 12 horas adelante. Las ANNs han superado el desempeño de la Persistencia y del Modelo Autoregresivo. Se ha demostrado que el empleo de las variables estacionales ha mejorado las predicciones de Kt realizadas por las ANNs.

Palabras clave: energía solar, predicciones, redes neuronales, índice de claridad, modelos autoregresivos.

INTRODUÇÃO

Tendo em vista a penetração crescente da energia fotovoltaica (fonte intermitente) no Sistema Elétrico de Potência (SEP) no Brasil, cresce a preocupação do Operador Nacional do Sistema (ONS) com o despacho das fontes em geral, posto que a intermitência (em escala horária) na saída das centrais fotovoltaicas pode acarretar dificuldades para a programação de fontes como, por exemplo, a fonte térmica (Sousa, 2013).

Com respeito à intermitência, existem basicamente três maneiras de se atacar o problema: a primeira abordagem é o armazenamento de energia (Liu et al., 2013), que se mostra como uma promessa para o futuro, porém economicamente inviável para a realidade atual; A segunda abordagem trata de tentar suavizar o problema através da integração geográfica das fontes baseada no efeito chamado de *spatial smoothing* (Ue et al., 2011; Hoff e Perez, 2010). A terceira abordagem baseia-se em previsões confiáveis da produção das centrais intermitentes em curto e médio prazos. Atualmente, esta abordagem apresenta-se como a solução mais imediata para o problema (Costa et al., 2008; Martín et al., 2010; Giebel et al., 2011), sendo, portanto, a abordagem adotada neste trabalho.

Dentre os fatores ambientais que influenciam na potência de saída de uma central fotovoltaica, a irradiância solar incidente é o fator preponderante. Uma das estratégias bastante utilizadas para auxiliar a previsão de irradiância consiste na transformação da série de irradiância em série de índice de claridade - Kt (McCandless et al., 2014). Com essa transformação, é possível remover grande parte das variações sazonais da radiação (Siqueira, 2008). A remoção da sazonalidade ajuda a tornar a série

estacionária, o que é um requisito para a aplicação de alguns modelos estatísticos, como, por exemplo, o modelo autorregressivo. Além disso, em redes neurais que utilizem como entradas variáveis de ordens de grandeza diferentes entre si recomenda-se que estas sejam normalizadas, o que já é característico do índice de claridade.

Diversos trabalhos foram publicados com respeito à previsão de radiação solar. Os trabalhos de Diagne et al. (2013), de Voyant et al. (2017) e de Inman et al. (2013) fazem uma profunda revisão dos modelos mais empregados nessa área. Martín et al. (2010) classificam-nos, de forma sucinta, como: a) modelos físicos; b) modelos estatísticos. Os modelos físicos são baseados em equações dinâmicas que descrevem o comportamento da atmosfera baseando-se em leis de conservação. Tais equações não possuem solução única devido às suas não linearidades. Por isso, recorre-se às soluções por aproximação numérica (Martín et al., 2010). São os chamados *Numerical Weather Prediction* (NWP) models. Os erros destes modelos variam significativamente e dependem, entre outros aspectos, do clima e dinâmica da radiação solar no local em estudo (Lorenz et al., 2009). A respeito da acurácia dos NWPs, Martín et al. (2010) indicam que, para um caso particular, os erros médios para a previsão horária de GHI (*Global Horizontal Irradiance*) variam entre 20,8% e 31,7% para as primeiras 24 horas e de 21,3% a 36,8% no segundo dia. Lima (2015) avaliou o desempenho de alguns modelos de previsão de irradiância no Nordeste do Brasil, encontrando um RMSE de 40% nas previsões de GHI horária do modelo numérico WRF.

Os modelos estatísticos são baseados na construção de relações empíricas entre preditores e preditando. Tais modelos (estatísticos) subdividem-se ainda em: modelos baseados em séries temporais (Box, 1976), modelos baseados em imagens de satélite (Perez et al., 2010) e modelos de downscaling, empregados para aumentar a resolução espacial da saída dos modelos físicos.

Em relação aos modelos baseados em séries temporais, diversas abordagens vêm sendo estudadas de forma a alcançar resultados mais acurados. As abordagens mais recorrentes nesse sentido são os modelos autorregressivos, as Redes Neurais e os modelos baseados em Lógica Difusa, conforme enfatizados nos trabalhos de Costa (2005) e de Diagne et al. (2013).

Os modelos autorregressivos são constituídos por uma abordagem fundada em análise de séries temporais através da extrapolação a um futuro próximo por meio de uma regressão linear utilizando uma estreita janela (composta pelos valores do presente e do passado recente). Esta janela é então deslocada ao longo de toda a série temporal com o objetivo de que os parâmetros da regressão linear sejam ajustados de forma adaptativa de modo a seguir a dinâmica da variável amostrada (Costa, 2005). A classe de métodos autorregressivos abrange as variantes AR (Autorregressivo), ARX (Autorregressivo com entradas exógenas), MA (Média Móvel), ARMA (junção do AR e MA), ARIMA (Autorregressivo Integrado de Médias Móveis, aplicáveis em séries temporais com evidência de não estacionariedade), entre outras (Box et al., 2015).

Redes Neurais são modelos matemáticos inspirados na estrutura neural humana e que adquirem conhecimento através da experiência. Por possuírem característica não linear, são muitas vezes aplicadas com sucesso na solução de problemas que englobam relações não lineares complexas entre as variáveis.

Mellit e Pavan (2010) estudaram previsão de irradiância solar em médio prazo (24h) direcionadas a plantas de geração fotovoltaicas conectadas à rede. O modelo proposto foi baseado em ANN e utilizava como entradas a irradiância média diária, a temperatura do ar média diária e o dia do mês. Já as saídas são irradiâncias (W/m^2) para um dia à frente com *time-step* de hora a hora. Os resultados do modelo atingiram coeficientes de correlação com dados medidos na faixa de 98-99% para dias de céu claro e de 94-96% para dias de céu nublado. Os erros RMSE da radiação solar prevista para dias de céu claro variaram de 18% a 67%. Já em dias de céu nublado, o erro atingiu variações de 67% a 85%.

Voyant et al. (2014) realizaram um estudo sobre previsão de radiação solar [Wh/m^2] utilizando como entrada 13 anos de radiação global no plano horizontal na região de Córsega, França. Os dados foram transformados em Índices de Céu Claro (K_{cs}), de forma conseguir uma nova série estacionária. As

previsões resultantes de modelos ARMA e ANN Perceptron Multicamada (MLP) atingiram nRMSE semelhantes, da ordem de 40% para um horizonte de 24h. Foi proposto um modelo híbrido, que alcançou um nRMSE de 36%.

Aguiar et al. (2016) propuseram um modelo de previsão de GHI horária para horizontes de 1h a 6h à frente utilizando dados medidos em terra (Ilhas Canárias, Espanha) e dados exógenos provindo de satélite (GHI) e previsões de Índice de Cobertura de Nuvens e radiação solar provindos do ECMWF (*European Centre for Medium-Range Weather Forecasts*). Foram realizadas comparações entre os modelos Persistência, ANN baseada em observações (modelo chamado de NN), saídas do ECMWF e dados de satélite (chamados de SAT), como também diversas combinações desses modelos. O melhor resultado foi alcançado pelo modelo combinado NN+SAT+ECMWF (rRMSE=24% para h=1h e rRMSE=34% para h=6h), o qual obteve uma melhoria máxima (h=6h à frente) de 3,5% em relação ao modelo NN, o que não é muito significativo tendo em vista o aumento da complexidade do modelo.

Este trabalho tem por objetivo a previsão de Kt com horizonte de previsão de 12h à frente e com intervalos de tempo horário. Foram testados modelos de previsão baseados em redes neurais, Autorregressivos e Persistência. Foram ainda introduzidas variáveis sazonais determinísticas a fim de melhorar o desempenho das redes neurais.

METODOLOGIA

A base de dados utilizada é composta por três anos (2013, 2014 e 2015) de séries observacionais provenientes da Estação SONDA Petrolina, latitude 09° 04' 08" S e longitude 40° 19' 11" O, localizada em uma região semiárida do Nordeste brasileiro. Trata-se de uma estação solarimétrica e anemométrica que dispõe de medição de diversas grandezas, dentre as quais Irradiância Global horizontal (GHI), Direta Normal (DNI) e difusa (Diff) minuto a minuto.

Os dados passaram por uma garantia de qualidade que utiliza diversos procedimentos retirados e adaptados de trabalhos anteriores (Younes et al., 2005; Long e Shi, 2006; Journée e Bertrand, 2011; Raichijk, 2012). Por conveniência, os dados de irradiância e de Kt foram passados da base minuto-a-minuto para a base horária. Além disso, foram utilizados apenas dados obtidos em ângulos de elevação solar $\alpha > 10^\circ$. Os demais dados foram retirados da série temporal. Após a etapa de garantia de qualidade, a base de dados foi dividida em dois subconjuntos: o de calibração dos modelos, formado pelos anos de 2013 e 2014, e o de validação dos modelos, formado pelos dados de 2015.

O trabalho buscou treinar duas categorias de modelos de regressão: os Modelos Autorregressivos (modelos lineares) e as ANN (modelos não lineares) de forma a testar qual deles se adaptaria melhor à previsão de Kt.

Modelos Autorregressivos (AR)

A seleção da ordem do modelo AR foi realizada baseada na análise da função de autocorrelação parcial da série de Kt, a qual indica quais *lags k* carregam em si uma quantidade de correlação significativa entre a amostra no instante de tempo t e a amostra no instante de tempo t-k. Realizada a seleção do modelo, os parâmetros do modelo AR foram então estimados através da regressão baseada no método dos mínimos quadrados. Para cada horizonte de tempo foi calibrado um modelo AR.

Modelos ANN

Assim como no caso dos modelos AR, foi calibrado um modelo ANN para cada horizonte de previsão utilizando o conjunto de dados de calibração. A estrutura de ANN utilizada neste trabalho foi a Perceptron Multicamada (MLP), a qual tem sido aplicada com sucesso em estimação de irradiância solar (Mellit e Pavan, 2010; Voyant et al., 2017). A função de ativação adotada foi a Tangente-Sigmóide e o algoritmo de aprendizagem para treinamento da rede foi o de Levenberg-Marquardt. As redes testadas possuem estruturas com até três camadas ocultas e, em cada uma delas, até oito neurônios, respeitando a característica triangular. Para a tarefa de treinar as várias combinações de ANN, foi utilizado o cluster Papa-léguas de alto desempenho pertencente ao CER-UFPE, com 22 CPUs AMD Opteron 6000 series, 352 cores e 384 Gb de RAM compartilhada. Em princípio, foram

usados como entrada da ANN apenas os dados de Kt. Tais modelos foram denominados de Modelos K.

É comum que, além da variável endógena, as variáveis data e hora sejam utilizadas como entradas de modelos baseados em ANN. Segundo Smith et al. (2006), a inserção dessas duas variáveis não é suficiente para que a ANN capte satisfatoriamente as informações de sazonalidade da variável em estudo. Para contornar tal problema, cada uma dessas variáveis (data e hora) foi dividida em quatro outras variáveis, totalizando oito, as quais são construídas utilizando conceitos de lógica difusa e com amplitudes variando entre 0 e 1. Estas 8 variáveis foram incluídas como entradas da rede, sendo 4 representando a sazonalidade das estações (Primavera, Verão, Outono e Inverno, ilustradas na Figura 1) e 4 representando a sazonalidade diária (Manhã, Meio-Dia, Tarde e Meia-Noite). A este novo modelo de ANN, contendo como entradas Kt e as variáveis sazonais, foi dado o nome de Modelo KS.

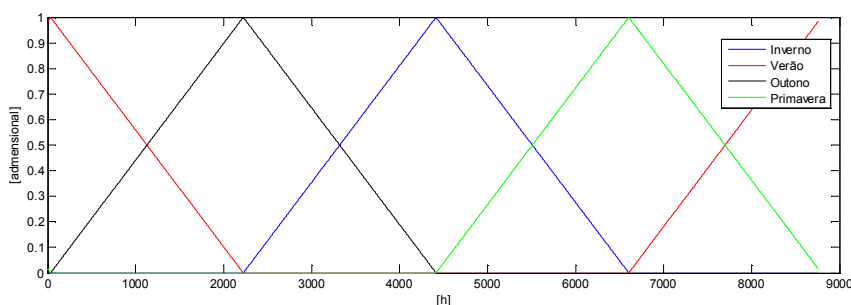


Figura 1: Variáveis Sazonais relativas às estações do ano, utilizadas no trabalho de Smith et al. (2006).

ANÁLISE DOS RESULTADOS

Para se avaliar o desempenho de um modelo de previsão, é comum entre muitos autores (Costa, 2005; Montgomery et al., 2012) a utilização de um modelo de referência, para assim comparar os resultados de ambos os modelos com as observações. O modelo de referência comumente utilizado é o Modelo de Persistência, devido a sua simplicidade na implementação, o qual é definido como:

$$\hat{y}_{t+k} = \hat{y}_t \quad (1)$$

onde \hat{y}_{t+k} representa a previsão k passos à frente. A comparação por persistência se justifica pelo fato de quando se compara, por exemplo, o RMSE encontrado por um autor com o RMSE encontrado nos trabalhos de outro autor, tal comparação não é adequada, devido aos experimentos terem sido realizados sob condições (local, época do ano, altitude, variabilidade do recurso, entre outras) diferentes (Inman et al., 2013).

Além da persistência, algumas métricas estatísticas costumam ser utilizadas para análise de desempenho dos modelos, as quais estão descritas na Tabela 1. Por fim, quando diversas metodologias são aplicadas sob as mesmas condições, ou seja, sob a mesma série observacional, convém realizar a comparação de desempenho dos mesmos através do Diagrama de Taylor (2001), o qual vem sendo amplamente utilizado pela comunidade científica. Trata-se de um diagrama que provê um resumo estatístico de quão bem os dados se equivalem entre si em termos de sua correlação, de seu RMSD e a amplitude de suas variâncias (representado por seus desvios padrão) (Taylor, 2001). Estes diagramas são especialmente úteis na avaliação de múltiplos aspectos de modelos complexos ou para comparar simultaneamente a habilidade relativa dos diversos modelos.

ESTATÍSTICO	DESCRIÇÃO
$\sigma_x = \left[\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right]^{1/2}$	Desvio Padrão de uma variável x qualquer, onde \bar{x} é a média de x.

$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2 \right]^{1/2}$	Raiz do Erro Quadrático Médio entre o valor estimado (previsto) \hat{x} e o valor observado x .
$nRMSE = \frac{RMSE}{\bar{x}} \times 100$	RMSE normalizado, onde \bar{x} é a média de x .
$RMSD = \left[\frac{1}{N} \sum_{i=1}^N [(\hat{x}_i - \bar{\hat{x}}) - (x_i - \bar{x})]^2 \right]^{1/2}$	Raiz do Erro Quadrático Médio dos desvios. $\bar{\hat{x}}$ representa o valor médio da série prevista.
$nRMSD = \frac{RMSD}{\bar{x}} \times 100$	RMSD normalizado, onde \bar{x} é a média de x .
$R = \frac{\frac{1}{N} \sum_{i=1}^N (x_i + \bar{x})(y_i + \bar{y})}{\sigma_x \sigma_y}$	Coefficiente de Correlação entre as variáveis x e y .
$FS = 1 - \frac{RMSE_m}{RMSE_p}$	<i>Forecasting Skill (Improvement)</i> , onde os índices m e p são relativos ao modelo em questão e à persistência, respectivamente.

Tabela 1: Estatísticos utilizados para auxiliar a comparação entre modelos.

Modelos Autorregressivos (AR)

Para selecionar a ordem do modelo AR, foi calculada a autocorrelação parcial da série de K_t , donde se observou que os lags com maior significância foram os dois primeiros. Dessa forma, foi escolhida a ordem $p=2$ para o modelo autorregressivo, em conformidade com baixas ordens descritas em outros trabalhos sobre irradiância solar (McCandless et al., 2014; Voyant et al., 2014). O aumento da ordem p do $AR(p)$ não traria nenhum ganho significativo suficiente para justificá-lo, conforme observado na Figura 2.

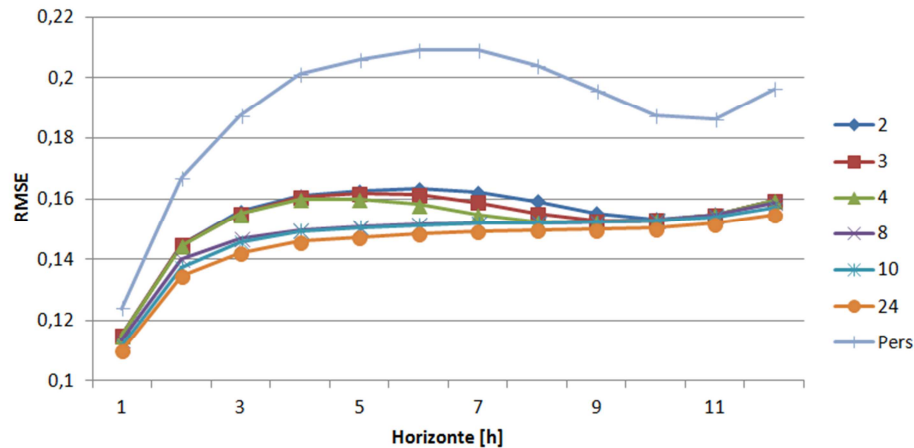


Figura 2: RMSE das previsões de K_t para diversas ordens p (2, 3, 4, 8, 10 e 24) dos modelos $AR(p)$ e para horizontes de previsão variando de uma a doze horas à frente.

Os erros para as previsões foram considerados altos, já que o RMSE se manteve na média em torno de 0,15. Tomando o K_t médio da série (0,55), tem-se um erro percentual $nRMSE$ próximo de 27%. Se for considerado apenas o horizonte uma hora à frente, o $nRMSE$ foi de 20,9%. Já o $nRMSD$ atingiu os 20%.

ANN – Modelo K.

Foram testados modelos com a ordem p variando de 1 a 8, o qual utiliza como entradas apenas os p valores de K_t (valor presente e “ $p-1$ ” valores passados). Por simplificação, serão mostrados apenas os

resultados para o horizonte 1 hora à frente. Para tal horizonte, a ANN que obteve melhor resultado possuía ordem $p=8$ e uma configuração de camadas ocultas [5 3], a qual possui duas camadas ocultas contendo cinco e três neurônios em cada uma, respectivamente. Tal modelo obteve um nRMSD=20,78%. A Figura 3 traz a comparação via diagrama de Taylor entre o Modelo K e a Persistência.

ANN – Modelo KS

Com o intuito de melhorar a acurácia do Modelo K, foram treinadas ANN que usaram como entradas, além de K_t , oito variáveis sazonais.

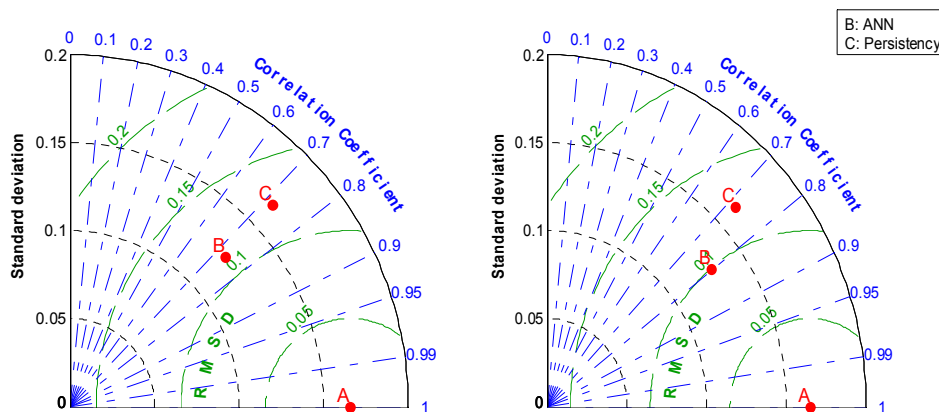


Figura 3: Diagrama de Taylor contendo a representação dos modelos K (letra B do diagrama à esquerda), KS (letra B do diagrama à direita) e Persistência (letra C) para uma hora à frente.

Pode-se perceber na Figura 3 que a inserção das variáveis sazonais provocou uma melhoria no desempenho do modelo baseado em ANN, onde houve uma diminuição no valor do RMSD e um aumento na correlação da saída do modelo em relação às observações. Enquanto que, para o horizonte de 1 hora à frente, o modelo AR obteve um nRMSD de 20%, o Modelo K e o Modelo KS obtiveram, respectivamente, nRMSD de 20,78% e 17,89%. O desempenho do Modelo KS para os demais horizontes de previsão podem ser observados na Figura 4.

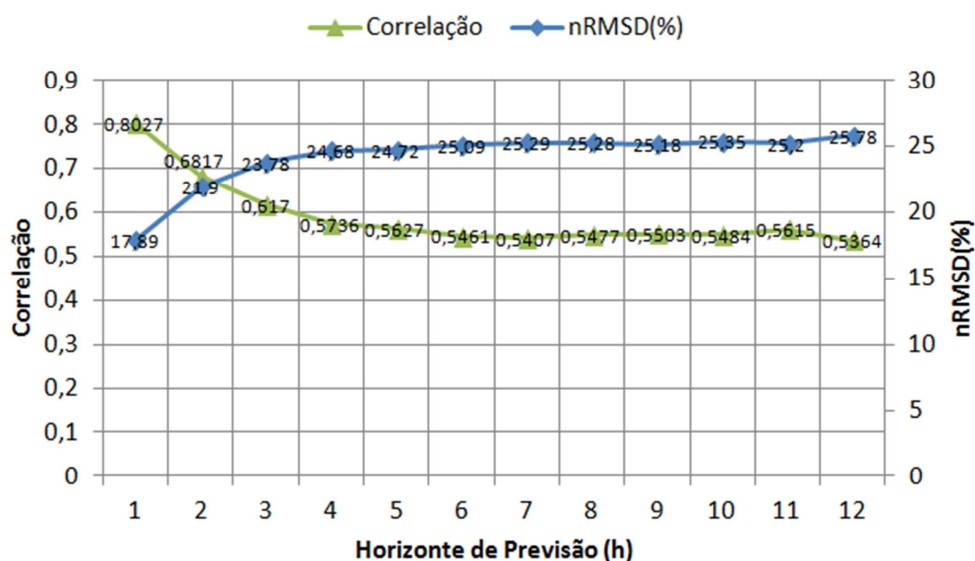


Figura 4: Parâmetros estatísticos das previsões do Modelo KS para os horizontes de 1h a 12h à frente.

Uma comparação entre os modelos Ar, KS e Persistência para os 12 horizontes de previsão pode ser observado na Figura 5.

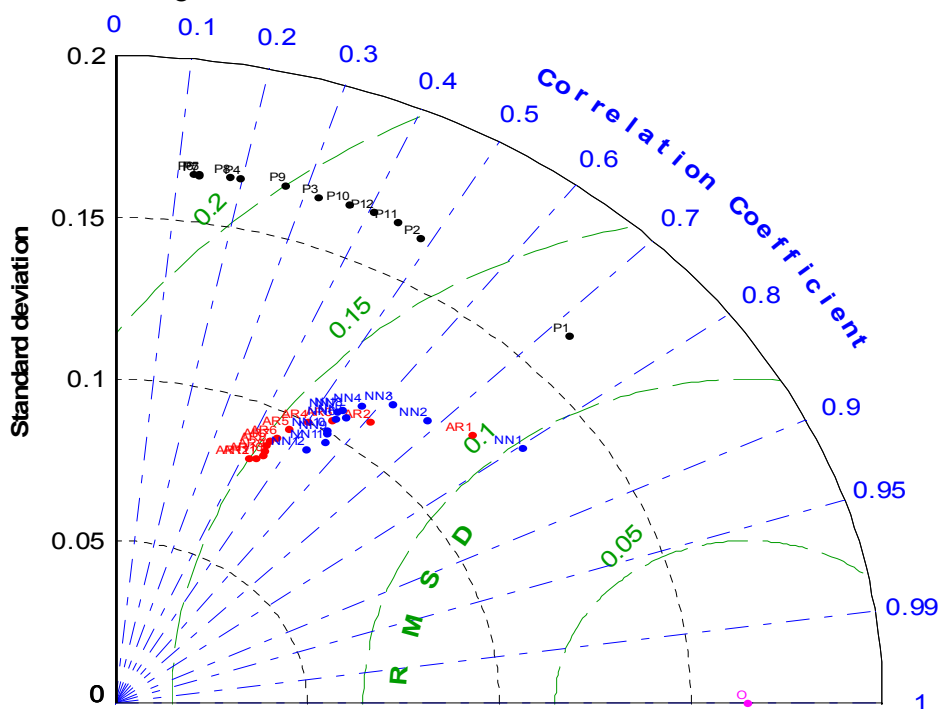


Figura 5: Comparação dos resultados das previsões dos modelos KS (representado por NN), Autorregressivo (AR) e Persistência(P) para horizontes de previsão de 1h a 12h à frente.

Quando se aumenta o horizonte de previsão, é esperada uma degradação nos resultados da previsão. A Figura 4, leva a concluir (erroneamente) que, a partir do horizonte 4, as previsões possuem a mesma acurácia, ou seja, praticamente não se degradaram. Porém, o diagrama de Taylor apresentado na Figura 5 nos permite observar uma leve degradação com o aumento do horizonte, principalmente no que diz respeito à diminuição da variabilidade (e conseqüentemente o desvio padrão) da série prevista. Este exemplo evidencia a importância de se utilizar tal diagrama quando se quer comparar simultaneamente a acurácia de vários modelos.

Pode-se concluir que as redes neurais superaram em todos os horizontes os modelos Autorregressivos. O modelo de Persistência, embora com variabilidade igual às observações (como esperado), foi superado tanto pelo autorregressivo quanto pelas ANN, conforme se encontra quantificado na Figura 6 e na Tabela 3, as quais também possibilitam a comparação dos modelos AR e KS com modelos utilizados em outros trabalhos (descritos na Tabela 2).

Nome do Modelo	Autor	Local	Descrição
Satélite	(Perez et al., 2010)	USA	Baseado em movimento das nuvens derivado de imagens de satélite
ARIMA Reikard	(Reikard, 2009)	Clark, Nevada, USA	Baseado em ARIMA e dados observacionais (GHI)
NN Aguiar	(Aguiar et al., 2016)	Ilhas Canárias, Espanha	Baseado em ANN e dados observacionais (GHI)
Híbrido Aguiar	(Aguiar et al., 2016)	Ilhas Canárias, Espanha	Baseado em ANN, GHI, ECMWF e dados de Satélite

Tabela 2: Modelos de previsão horária de GHI de outros autores que serão comparados aos modelos testados no presente trabalho.

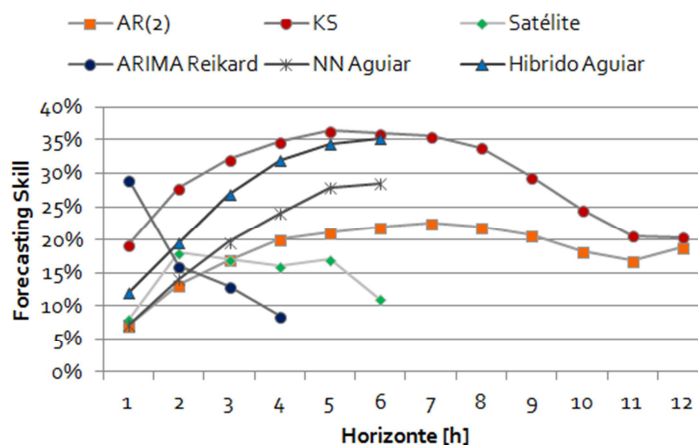


Figura 6: Melhoria (Forecasting Skill) dos modelos AR(2) e KS, como também de modelos utilizados em outros trabalhos.

Modelo	Horizonte à frente [h]											
	1	2	3	4	5	6	7	8	9	10	11	12
AR(2)	7,1%	13,1%	17,0%	20,0%	21,1%	22,0%	22,6%	22,0%	20,7%	18,3%	16,9%	18,8%
KS	19,3%	27,9%	32,1%	34,7%	36,3%	36,0%	35,5%	33,8%	29,5%	24,6%	20,6%	20,4%
Satélite	8,0%	18,0%	17,0%	16,0%	17,0%	11,0%	-	-	-	-	-	-
ARIMA Reikard	29,0%	16,0%	13,0%	8,5%	-	-	-	-	-	-	-	-
NN Aguiar	7,2%	14,0%	19,6%	24,0%	27,8%	28,5%	-	-	-	-	-	-
Híbrido Aguiar	11,9%	19,6%	26,8%	31,9%	34,4%	35,2%	-	-	-	-	-	-

Tabela 3: Valores percentuais da melhoria para os diversos horizontes de previsão dos modelos AR e KS e de modelos utilizados por outros autores.

Analisando a Figura 6, percebe-se que o modelo KS supera em todos os horizontes todos os modelos com os quais foi comparado, com exceção do modelo ANN Reikard, que supera no primeiro horizonte (apenas) o modelo KS. Isso pode ser explicado pelo fato do modelo ANN Reikard ser um modelo concebido para o curto prazo. Além disso, o modelo KS (univariado) superou modelos mais complexos que utilizam variáveis exógenas, como o modelo Híbrido Aguiar.

CONCLUSÕES

Este trabalho se propôs a abordar uma melhoria dos modelos baseados em ANN de previsão de Kt através de uma prática antes aplicada na previsão de temperatura (Smith et al., 2006). Essa melhoria consistiu na inclusão de variáveis sazonais pertencentes a um conjunto difuso. Outros modelos (autorregressivos e persistência) foram também utilizados para comparação de desempenho. As previsões foram realizadas para horizontes variando de 1h a 12h à frente.

A introdução das variáveis sazonais propiciou uma melhoria no desempenho das redes neurais para a previsão de séries de Kt, fazendo com que o nRMSD melhorasse de 20,78% para 17,89% para o primeiro horizonte de previsão. Além disso, a correlação passou de 0,73 para 0,8. A variabilidade da série prevista neste horizonte de previsão praticamente não se alterou.

O modelo KS superou os modelos AR e Persistência em todos os horizontes de previsão. Os resultados obtidos pelo Modelo KS, mesmo em horizontes mais altos, são considerados bons quando comparados a trabalhos de outros autores, como Aguiar et al. (2016) e Mellit e Pavan (2010). Como perspectivas de trabalhos futuros, a introdução de variáveis exógenas, tais como a temperatura, a umidade e a pressão, poderá vir a melhorar a acurácia das previsões dos modelos aqui estudados.

REFERENCIAS

- Aguiar, L. M., Pereira, B., Lauret, P., Díaz, F. e David, M. (2016). Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy* 97: 599-610.
- Box, G. E., Jenkins, G. M., Reinsel, G. C. e Ljung, G. M. (2015). *Time series analysis: forecasting and control*, John Wiley & Sons.
- Box, G. E. P. (1976). *Time series analysis : forecasting and control*. Oakland (Cal.), Holden-Day.
- Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H. e Feitosa, E. (2008). A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews* 12(6): 1725-1744.
- Costa, A. A. (2005). *Matemathical/statistical and physical/meteorological models for short-term prediction of wind farms output*, Universidad Politécnica de Madrid.
- Diagne, M., David, M., Lauret, P., Boland, J. e Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews* 27: 65-76.
- Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M. e Draxl, C. (2011). The state-of-the-art in short-term prediction of wind power: A literature overview.
- Hoff, T. E. e Perez, R. (2010). Quantifying PV power Output Variability. *Solar Energy* 84(10): 1782-1793.
- Inman, R. H., Pedro, H. T. C. e Coimbra, C. F. M. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science* 39(6): 535-576.
- Journée, M. e Bertrand, C. (2011). Quality control of solar radiation data within the RMIB solar measurements network. *Solar Energy* 85(1): 72-86.
- Lima, F. J. L. (2015). *Previsão de Irradiação Solar no Nordeste do Brasil Empregando o Modelo WRF Ajustado por Redes Neurais Artificiais(RNAs)*. Doutorado.
- Liu, J., Zhang, J. G., Yang, Z., Lemmon, J. P., Imhoff, C., Graff, G. L., Li, L., Hu, J., Wang, C. e Xiao, J. (2013). Materials science and materials chemistry for large scale electrochemical energy storage: from transportation to electrical grid. *Advanced Functional Materials* 23(8): 929-946.
- Long, C. N. e Shi, Y. (2006). *The QCRad value added product: Surface radiation measurement quality control testing, including climatology configurable limits*. Atmospheric Radiation Measurement Program Technical Report.
- Lorenz, E., Hurka, J., Heinemann, D. e Beyer, H. G. (2009). Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of selected topics in applied earth observations and remote sensing* 2(1): 2-10.
- Martín, L., Zarzalejo, L. F., Polo, J., Navarro, A., Marchante, R. e Cony, M. (2010). Prediction of global solar irradiance based on time series analysis: application to solar thermal power plants energy production planning. *Solar Energy* 84(10): 1772-1781.
- McCandless, T. C., Haupt, S. E. e Young, G. S. (2014). *Short Term Solar Radiation Forecasts Using Weather Regime-Dependent J3. 5 Artificial Intelligence Techniques*.
- Mellit, A. e Pavan, A. M. (2010). 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* 84(5): 807-821.
- Montgomery, D. C., Peck, E. A. e Vining, G. G. (2012). *Introduction to linear regression analysis*, John Wiley & Sons.
- Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renné, D. e Hoff, T. E. (2010). Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy* 84(12): 2161-2172.
- Raichijk, C. (2012). *Control de calidad de mediciones de radiación solar*. Presentado en XXXV Reunión de Trabajo de la Asociación Argentina de Energías Renovables y Ambiente (ASADES), Rosario, Argentina.
- Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy* 83(3): 342-349.
- Siqueira, A. N. (2008). *Geração sintética da irradiação solar diária no Brasil para aplicações energéticas*.

- Smith, B. A., McClendon, R. W. e Hoogenboom, G. (2006). Improving air temperature prediction with artificial neural networks. *International Journal of Computational Intelligence* 3(3): 179-186.
- Sousa, J. (2013). A Importância da Previsão da Geração Elétrica de Base Eólica. Brazil Windpower 2013 Conference and Exhibition. Brazil.
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* (1984–2012) 106(D7): 7183-7192.
- Ue, Y., Hara, R., Kita, H., Saito, Y., Takitani, K. e Saito, M. (2011). Analysis of Spatial Smoothing Effect of Short-term Fluctuation of Global Solar Radiation based on the Data Including Quantization Error. *IEEJ Transactions on Power and Energy* 131: 413-420.
- Voyant, C., Darras, C., Muselli, M., Paoli, C., Nivet, M.-L. e Poggi, P. (2014). Bayesian rules and stochastic models for high accuracy prediction of solar radiation. *Applied Energy* 114: 218-226.
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F. e Foulloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy* 105: 569-582.
- Younes, S., Claywell, R. e Muneer, T. (2005). Quality control of solar radiation data: present status and proposed new approaches. *Energy* 30(9): 1533-1549.

ABSTRACT

This work aims to improve solar radiation forecast for horizons up to 12 hours ahead. In this sense, Neural Networks (ANN) of Multilayer Perceptron type were proposed, using as regressive variables, hourly clearness index (Kt) and seasonal deterministic variables, with the purpose of enhancing the understanding on the characteristics of daily and annual seasonal of Kt series for the city of Petrolina-PE-Brazil. The objective is to compare the results obtained with the use of such seasonal variables (which, among other aspects, have characteristics of fuzzy variables) with results of other models reported in the literature, such as ANNs without the use of seasonal variables, the Auto-Regressive Model and Persistence. The error (nRMSD) associated with the final model differs from 17.8% to 25.7% for horizons from 1 to 12 hours ahead. Neural networks surpassed the Persistence and the Autoregressive Model. It was proved that the use of the seasonal variables improved the results of Kt predictions using ANNs.

Keywords: solar energy, forecasting, neural networks, clearness index, autoregressive models.