



Metodologias de correlação espaço temporal e sintetização de séries temporais de irradiação para previsão de recurso energético de centrais solares

Paulo André de Melo Ferreira Sampaio

VERSÃO DE TRABALHO

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Orientador: Professor Doutor Cláudio Monteiro

Orientador na empresa: Eng. Paulo Pinto

25 de Julho de 2019

Resumo

Nesta dissertação recorre-se ao desenvolvimento de metodologias e programação de extração de informação e síntese de séries temporais com o objetivo de gerar um ano meteorológico típico representativo do recurso energético das centrais solares em estudo. Para isso, utilizam-se modelos baseados em Redes Neurais (RN) tendo em conta valores de Longo Termo obtidos com base em dados históricos de radiação solar e previsões meteorológicas, bem como valores medidos localmente de um caso de estudo exemplo fornecido pela empresa Megajoule.

Estes modelos pretendem definir a melhor correlação de irradiação e temperatura em relação aos dois tipos de registos fornecidos, registos de Longo Termo e registos medidos localmente.

As redes com melhores desempenhos dos modelos de Redes Neurais propostos são aplicadas em diferentes conjuntos de teste de forma a avaliar-se a qualidade da correlação encontrada.

Na fase final são construídos dois conjuntos de TMY (ano meteorológico típico) baseados em Redes Neurais. Por fim, os TMY criados, o TMY resultante da aplicação da metodologia utilizada pela Megajoule e o TMY inicial de Longo Termo fornecido por fontes meteorológicas são comparados com registos de temperatura e irradiação de valores horários medidos do ano de 2017 e da média horária dos anos de 2014, 2015 e 2017.

Concluiu-se, por fim, que os TMY criados através de Redes Neurais apresentam melhorias de correlação em relação aos registos de medições locais disponíveis, por comparação com os restantes TMY. Os TMY criados são, portanto, uma alternativa válida de MCP (medição-correlação-previsão) solar.

Abstract

This dissertation is based on the development of methodologies and programming of data extraction and synthesis of time series with the objective of generating a meteorological model representative of the energy resource of the solar arrays under study. For this, models based on Artificial Neural Networks (ANN) are used taking into account Long Term values obtained based on historical solar radiation and meteorological data, as well as values measured locally from an example case study provided by Megajoule.

These models aim to define the best correlation between irradiation and temperature in relation to the two types of records provided, Long Term records and locally measured records.

The best performing networks of the proposed Artificial Neural Networks models are applied in different test sets in order to evaluate the quality of the correlation.

In the final phase two sets of TMY (Typical Meteorological Year) are built based on Artificial Neural Networks. Finally, the TMY created, the TMY resulting from the application of the methodology used by Megajoule and the initial Long Term TMY provided by weather sources are compared with temperature and irradiance records of measured time values of the year 2017 and the hourly average of the years 2014, 2015 and 2017.

Finally, it was concluded that the TMYs created through Artificial Neural Networks show correlation improvements over the available local measurement records compared to the remaining TMYs. Artificial Neural Networks based TMYs therefore represent a valid alternative to solar MCP (measurement-correlation-prediction).

Agradecimentos

A conclusão desta dissertação marca o final de uma etapa de grandes lutas e conquistas e sobretudo de grandes amizades que ficam neste término da vida académica. Deste modo gostaria de expressar os meus mais sinceros agradecimentos a todos aqueles que fizeram parte desta caminhada.

Em primeiro lugar, agradeço à minha família, em particular aos meus pais e à minha irmã, por todo o suporte e encorajamento que me deram ao longo de todos estes anos.

De seguida, agradeço ao Professor Doutor Cláudio Monteiro, pela orientação, disponibilidade e conselhos fornecidos ao longo de todo o desenvolvimento desta dissertação.

Gostaria de agradecer também ao Engenheiro Paulo Pinto, pela oportunidade de realizar esta dissertação num contexto empresarial e pela disponibilidade e aconselhamento transmitidos ao longo da dissertação. Uma palavra também a todos os colegas dentro da empresa Megajoule que se mostraram sempre disponíveis para ajudar. Um agradecimento especial para a Eng. Catarina Novais por todo o apoio fornecido.

Por último, um agradecimento a todos os amigos que partilharam comigo este percurso.

Paulo André de Melo Ferreira Sampaio

“Intelligence without ambition is a bird without wings.”

Salvador Dali

Conteúdo

1	Introdução	1
1.1	Contextualização e motivação	1
1.2	Objetivos	3
1.3	Estrutura da Dissertação	3
1.4	Dados utilizados	4
1.5	Contexto da dissertação dentro da empresa MegaJoule	4
2	Estado da arte	7
2.1	Radiação Solar	7
2.1.1	Componentes da radiação	8
2.1.2	Posição do Sol: coordenadas horárias	10
2.1.3	Posição solar:coordenadas angulares	12
2.1.4	Cálculo do index de claridade	13
2.2	Estimativa do recurso solar	14
2.2.1	Modelos de previsão solar	14
2.2.2	Previsão de radiação solar utilizando imagens obtidas por satélite (sky cameras)	17
2.2.3	Modelos de previsão numérica de tempo (NWP)	18
2.3	Redes Neurais Artificiais (ANN)	19
2.3.1	Neurónios	20
2.3.2	Arquitetura da rede	21
2.3.3	Processo de aprendizagem	21
2.3.4	Normalização de variáveis	22
2.3.5	Avaliação do desempenho dos modelos de previsão	23
2.4	Métodos de MCP solar	24
2.4.1	Métodos utilizados na empresa Megajoule	24
3	Processamento e Análise de dados	27
3.1	Introdução	27
3.2	Caracterização dos dados utilizados	28
3.2.1	Dados de Longo Termo	28
3.2.2	Dados de TMY	29
3.2.3	Dados medidos	30
3.2.4	Dados de céu limpo (Clear sky)	31
3.3	Sincronização de dados	31
3.3.1	Filtragem e processamento dos dados de treino e de teste	32
3.4	Software Utilizado	34
3.4.1	Nftool	34

3.5	Construção da rede neuronal	35
3.5.1	Identificação de modelos de RN: variáveis a prever e potenciais entradas .	36
3.5.2	Normalização	38
3.5.3	Avaliação das variáveis de entrada da rede	39
3.5.4	Algoritmo desenvolvido para teste e treino de rede	40
3.5.5	Avaliação de redes	42
3.6	Aplicação de redes neurais	44
3.6.1	Comparação de métodos	44
3.6.2	Construção do TMY	44
3.7	Comparação do TMY	46
4	Exposição de resultados	49
4.1	Redes neurais	49
4.1.1	Modelo 1	49
4.1.2	Modelo 2	51
4.1.3	Modelo 3	52
4.1.4	Comparação modelo 1 e modelo 3	54
4.2	Comparação entre métodos	54
4.2.1	Construção do TMY	54
4.3	Comparação de TMY's com valores medidos de 2017	56
4.3.1	Irradiância	56
4.3.2	Temperatura	57
4.4	Comparação de TMY's com média da totalidade de valores medidos	59
4.4.1	Irradiância	59
4.4.2	Temperatura	60
5	Conclusões e Trabalho Futuro	63
5.1	Conclusões	63
5.2	Trabalho Futuro	64
	Referências	65

Lista de Figuras

1.1	Evolução da potência fotovoltaica acumulada instalada (GW)	2
1.2	Potência fotovoltaica instalada no ano de 2017 (GW)	2
2.1	Atlas da distribuição de irradiação direta normal.	7
2.2	Dependência da AM em relação à posição do observador.	9
2.3	Componentes da radiação solar resultantes da interação com a atmosfera terrestre.	10
2.4	Série de tempo de componentes de irradiância solar medidas por pireliómetro (A=DNI) e piranómetros (B=GHI);C=DHI)[1]	13
2.5	Imagem conceitual da eficiência de previsão em relação ao prazo de previsão previsto para os métodos baseados no solo, em satélite ou em NWP.	17
2.6	Método de vetores de deslocamento de nuvens	18
2.7	Estrutura de uma rede neuronal simples com 3 camadas	20
2.8	Modelo simplificado da estrutura de um neurónio	21
3.1	Comportamento das variáveis temperatura e irradiância nos registos de TMY	30
3.2	Clear sky, irradiância de longo termo e irradiância medida num dia de verão e num dia de inverno.	32
3.3	Dados utilizados nas redes neuronais para modelos de irradiância	33
3.4	Dados utilizados nas redes neuronais para modelos de temperatura	34
3.5	Rede de treino representativa das entradas da rede, camada escondida, camada de saída e saída da rede.	35
3.6	Treino dos 3 modelos: irradiância, temperatura e index de claridade	36
3.7	Fluxograma do algoritmo utilizado	42
3.8	Intersecção entre $F_1(x)$ e $1-F_2(x)$	43
4.1	Comparação das redes com melhores performances do modelo 1	50
4.2	Comparação das redes com melhores performances do modelo 2	52
4.3	Comparação das redes com melhores performances do modelo 3	53
4.4	Média mensal de irradiância dos dados de Longo termo correlacionados (W/m^2)	55
4.5	Desvios absolutos em relação à média para cada mês (W/m^2). Eixo dos yy's representa o ano cujo o desvio absoluto desse mês foi menor	55
4.6	Erro médio absoluto (MAE) mensal de irradiância em relação ao valores do ano de 2017	56
4.7	Ganho (%) em relação ao MAE mensal do modelo de referência (TMY inicial)	57
4.8	Erro médio absoluto (MAE) mensal de temperatura em relação ao valores do ano de 2017	58
4.9	Ganho (%) em relação ao MAE mensal do modelo de referência (TMY inicial)	58
4.10	Erro médio absoluto (MAE) mensal de irradiância em relação à média da totalidade dos valores medidos	59

4.11	Ganho (%) em relação ao MAE mensal do modelo de referência (TMY inicial) .	60
4.12	Erro médio absoluto (MAE) mensal de temperatura em relação à média da totalidade dos valores medidos	60
4.13	Ganho (%) em relação ao MAE mensal do modelo de referência (TMY inicial) .	61

Lista de Tabelas

3.1	Descrição de valores dos dados de Longo Termo	29
3.2	Anos escolhidos como "mais típicos" para cada mês do ano	29
3.3	Descrição de valores dos dados de TMY	30
3.4	Descrição de valores medidos	31
3.5	Coeficientes de correlação de Pearson entre as possíveis variáveis de entrada. . .	37
3.6	Coeficientes de correlação entre as possíveis variáveis de entrada e as variáveis de saída dos 3 modelos (R)	37
3.7	Exemplo de dados exportados após terminar o algoritmo	41
3.8	Exemplo da comparação de duas redes	43
4.1	Melhores redes encontradas com a utilização do <i>Forward Method</i> para o modelo 1. .	49
4.2	Melhores redes encontradas no <i>Backward method</i> para o modelo 1.	50
4.3	Aumento de neurónios das redes anteriores com melhores performances, N representa o número de neurónios	50
4.4	Melhores redes encontradas no <i>Forward method</i> para o modelo 2.	51
4.5	Melhores redes encontradas no <i>Backward method</i> para o modelo 2.	51
4.6	Aumento de neurónios das redes anteriores com melhores performances do modelo 2, N representa o número de neurónios	51
4.7	Melhores redes encontradas no <i>Forward method</i> para o modelo 3.	52
4.8	Melhores redes encontradas no <i>Backward method</i> para o modelo 3.	53
4.9	Aumento de neurónios das redes anteriores com melhores performances do modelo 3, N representa o número de neurónios	53
4.10	Resultados dos parâmetros utilizados para a comparação entre métodos de MCP .	54
4.11	Ano inserido no TMY const para cada mês e desvio absoluto em relação à média mensal de Longo Termo	56
4.12	Parâmetros de comparação de TMY's	57
4.13	Parâmetros de comparação de TMY's	59
4.14	Parâmetros de comparação de TMY's	60
4.15	Parâmetros de comparação de TMY's	61

Abreviaturas e Símbolos

Lista de Abreviaturas:

AM	Massa Relativa do ar
ANN	Artificial Neural Networks
CMV	Cloud motion vectors
CS	Clear sky
DHI	Irradiação difusa horizontal
DNI	Irradiação direta normal
ECMWF	European Centre for Medium-range Weather Forecasting
ET	Equação do tempo
FV	Fotovoltaico
GEFS	Global Ensemble Forecast System
GFS	Global Forecast System
GHI	Irradiação global horizontal
HO	Hora Oficial do País
IEA	International Energy Agency
LT	Longo Termo
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBE	Mean Bias Error
MCP	Measure Correlate Predict
MLP	Multilayer Perceptron
NAM	North American Mesoscale
NCEI	National Centers for Environmental Information
NWP	Numerical Weather Prediction
PVGIS	Photovoltaic Geographical Information System
RMSE	Root Mean Absolute Error
RN	Redes Neurais
SDE	Standard Deviation Error
TMY	Typical Meteorological Year
TSM	Tempo solar verdadeiro
TSV	Tempo solar verdadeiro

Lista de Símbolos:

G_{clear}	Irradiância de um dia de céu limpo
K_t	Index de Clareza
Kt_M	Index de clareza Medido
Kt_{LT}	Index de clareza de Longo Termo
I_{sc}	Constante solar
Imp_{MAE}^{ref}	Quantificador de ganho em relação a modelo de referência
O_h	Valores horários do subconjunto de teste
O_h^{RN}	Valores horários do subconjunto de saída das Redes Neurais
P_{meas}	Outputs medidos
P_{pred}	Outputs previstos
Rad_M	Irradiância Horizontal Global Medida
$Temp_M$	Temperatura ambiente Medida

Capítulo 1

Introdução

Neste capítulo inicial é efetuado um enquadramento em relação ao tema da dissertação , onde são definidos os objetivos e a estrutura inerente à sua realização.

Na parte final é apresentada uma descrição do contexto empresarial e é estabelecida uma referência à informação utilizada ao longo do trabalho.

1.1 Contextualização e motivação

O aquecimento global e a utilização de forma abusiva dos combustíveis fósseis ao longo das últimas décadas, criaram diversos problemas que têm justificado a procura de soluções alternativas, baseadas, por exemplo , no uso e desenvolvimento de fontes de energia renovável. A utilização de fontes de energia renovável, entre elas: a energia solar, eólica, hídrica e geotérmica, tem sido reconhecida como uma solução inovadora para o problema mencionado. [2]

Por substituição de fontes de energia convencionais, a energia solar fotovoltaica tem emergido como a solução mais popular, sendo a mais utilizada em muitos países por comparação com outras soluções. A energia solar transformou-se, assim , na fonte mais promissora de geração de energia a nível residencial, comercial e em aplicações industriais.[3] O número e o tamanho de parques e instalações fotovoltaicas cresceu de forma astronómica nos últimos anos. No final do ano de 2017, a potência global instalada chegou aos 403,3 GW, figura 1.1. Os países inseridos na *International Energy Agency* IEA representam 348,1 GW dessa instalação. Em adição, 37 países que não fazem parte do programa IEA representaram cerca de 55,2 GW de potência instalada.[4]

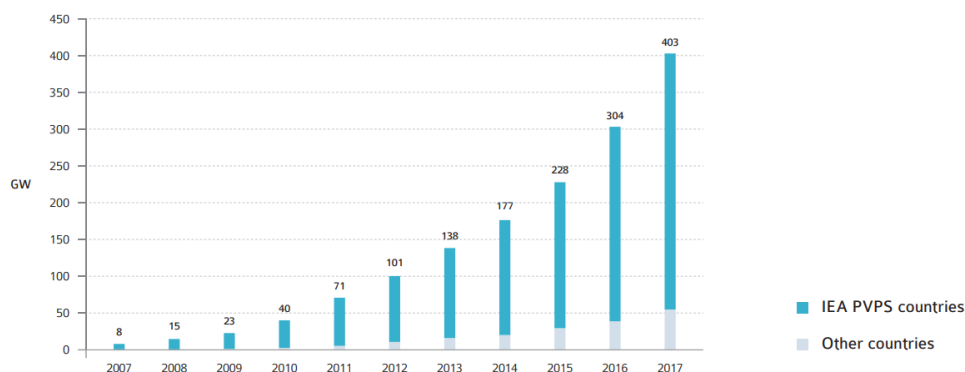


Figura 1.1: Evolução da potência fotovoltaica acumulada instalada (GW)

Os 27 países pertencentes ao IEA instalaram pelo menos 83,4 GW de potência no ano de 2017, enquanto que os países não pertencentes contribuíram com cerca de 15,5 GW. Com 403,3 GW de capacidade acumulada instalada, no ano de 2017 houve um crescimento de cerca de 33 %, recorde de capacidade instalada num ano, figura 1.2

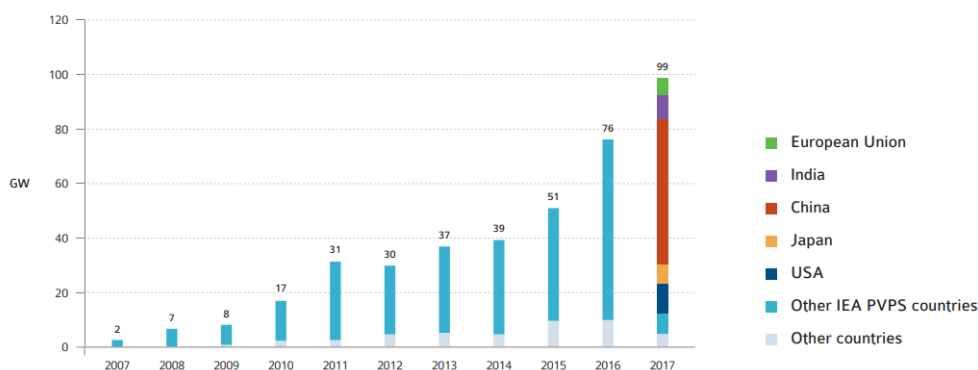


Figura 1.2: Potência fotovoltaica instalada no ano de 2017 (GW)

O aumento da construção de parques fotovoltaicos (FV) e a crescente implementação de sistemas interligados e isolados introduz mudanças dramáticas nos sistemas de energia que devem ser planeadas e antecipadas. A implementação em grande escala de instalações FV ligadas à rede implica problemas significantes relativos, por exemplo, à estabilidade do sistema, compensação de potência reativa e resposta da frequência. [5] Desta forma, a previsão de produção solar é um aspecto crucial para a modelização e planificação das instalações, previsões mais precisas permitem estimar reservas, gerir congestionamentos de rede, otimizar a gestão do armazenamento e reduzir custos de produção de eletricidade. [6]

1.2 Objetivos

Com vista a reduzir o período de medições e recolha de informação bem como aumentar a qualidade das previsões, serão desenvolvidos modelos de previsão multi-temporal para recriar séries temporais de histórico para o local da central em análise.

De forma sintetizada, os objetivos principais que deverão ser atingidos no desenvolvimento desta dissertação são os seguintes:

- Desenvolver uma metodologia que permita realizar uma correlação eficiente entre séries temporais de referência TMY para grandezas meteorológicas (irradiância, temperatura) e campanhas de medições locais, permitindo utilizar essa metodologia para recriar séries temporais para os locais em análise;
- Comparar a metodologia desenvolvida com as metodologias utilizadas na empresa Megajoule;
- Aplicar a metodologia e validar a correlação obtida de irradiância e temperatura com a totalidade das medições locais disponíveis.

1.3 Estrutura da Dissertação

Esta dissertação será dividida em 5 capítulos:

O capítulo 1, referente à introdução, destina-se à contextualização da presente dissertação, onde é explicada, de forma introdutória, a motivação e os principais objetivos inerentes à sua elaboração.

O capítulo 2, referente ao estado da arte, destina-se à pesquisa e apresentação das metodologias utilizadas na empresa Megajoule já adotadas como forma de resolução do problema proposto, bem como recolher informação necessária para a contextualização das etapas utilizadas na elaboração da nova metodologia.

No capítulo 3, referente à análise de dados e metodologia, procede-se à caracterização e processamento da informação utilizada no caso de estudo, nomeadamente no que se refere a sincronização de dados e possíveis normalizações. Além disso, este capítulo visa explicar e justificar as etapas propostas na metodologia elaborada, alicerçado por fórmulas matemáticas e fluxogramas.

No capítulo 4, referente à aplicação e validação da metodologia, procura-se expor os resultados obtidos na sua aplicação, estabelecendo comparações em relação a outras metodologias e analisando os resultados obtidos.

No capítulo 5, referente às conclusões e trabalho futuro, é elaborado um resumo das conclusões gerais mais relevantes obtidas na elaboração da dissertação e são sugeridos trabalhos futuros no mesmo âmbito.

1.4 Dados utilizados

Na realização desta dissertação foram utilizados 2 tipos de dados para o caso de estudo exemplo: dados de Longo Termo (LT) e dados de Medições locais.

Os dados de LT foram recolhidos pelo *software* PVGIS que providencia dados meteorológicos de bases de dados históricos obtidas por satélite. Neste caso, foram fornecidos dados horários dos anos de 2005 a 2016 de irradiância horizontal global, elevação solar e temperatura ambiente, bem como dados de *Typical Meteorological Year* (TMY), dados horários representativos de um ano, com informação relativa a irradiância horizontal global, irradiância normal direta, irradiância difusa, humidade relativa, temperatura ambiente, velocidade do vento e direção do vento.

Os dados de medições locais são dados medidos de 10 em 10 minutos de irradiância horizontal global e temperatura ambiente recolhidos, respetivamente, por piranómetros instalados na horizontal e sensores de temperatura, pela empresa Megajoule para os anos de 2014, 2015 e 2017.

1.5 Contexto da dissertação dentro da empresa MegaJoule

A empresa Megajoule presta serviços de consultoria para avaliação de recursos eólicos, solares e projetos de biomassa. A consultoria de energia solar engloba serviços de *Due diligence* para centrais de energia solar, prospeção de recursos solares e estudos de viabilidade.

- **Due diligence** — Fornecimento de serviços de análise e suporte, que englobam a possibilidade de substituição de equipamentos e montagens ou extensão/ampliação das centrais, permitindo aos investidores perceberem quais os riscos associados ao setor da energia solar;
- **Prospeção dos recursos solares** — A partir de informação recolhida através de satélites, dados provenientes de estações meteorológicas e/ou medições pontuais a Megajoule utiliza um algoritmo auto-construído para determinar os parâmetros de radiação solar para toda a área solicitada num determinado projeto. O mapeamento do recurso poderá englobar resoluções que variem entre 100 m e 10 km;
- **Estudos de viabilidade** — O principal objetivo passa por fornecer uma análise dos múltiplos parâmetros do local que possam viabilizar/inviabilizar o projeto de instalação duma central de energia solar: restrições locais de construção, formas de acesso ao local, tecnologias em uso, estimativas de produção, etc.

Na avaliação do potencial de centrais solares, utilizam-se séries de Longo Termo de irradiação recolhidas pelos satélites próximos do local em estudo, referentes a dados obtidos nos últimos anos, combinadas com medições locais. Com a informação recolhida dos satélites a descrição do clima local é efetuada através do chamado ano meteorológico típico (TMY). Por forma a ajustar o TMY, gerado a partir dos dados de Longo Termo, aos dados medidos recolhidos pela empresa, recorre-se a técnicas de correlação espaço temporal, com o objetivo de recriar séries temporais para os próximos anos, permitindo avaliar o potencial do projeto quanto às estimativas de produção.

Neste contexto, a presente dissertação destina-se ao desenvolvimento e criação de modelos de correlação e previsão multi-temporal, com o objetivo de desenvolver uma metodologia que permita ajustar o TMY inicial com maior precisão do que as metodologias de MCP (*measure-correlate-predict*) utilizadas.

Capítulo 2

Estado da arte

2.1 Radiação Solar

A radiação solar corresponde à energia fornecida pelo Sol [7]. Devido à grande distância entre o Sol e a Terra, aproximadamente $1,495 \times 10^8$ km, apenas uma pequena parte dessa energia atinge, efetivamente, a superfície terrestre. Dos $62,5\text{MW}/\text{m}^2$ de irradiância emitidos da superfície solar, apenas cerca de $1\text{kW}/\text{m}^2$ chega à superfície terrestre, valor denominado de constante solar [8]

As diferenças de radiação solar anuais e diárias nas diversas zonas ao longo da superfície terrestre provocam fenômenos meteorológicos (vento, chuva, neve, etc), fenômenos esses que resultam na formação de correntes e armazenamento de água (lagos, rios, marés, etc). Por outras palavras, é possível dizer que a maior parte das energias utilizadas, à exceção da energia nuclear e geotérmica, são provenientes do Sol.[9]

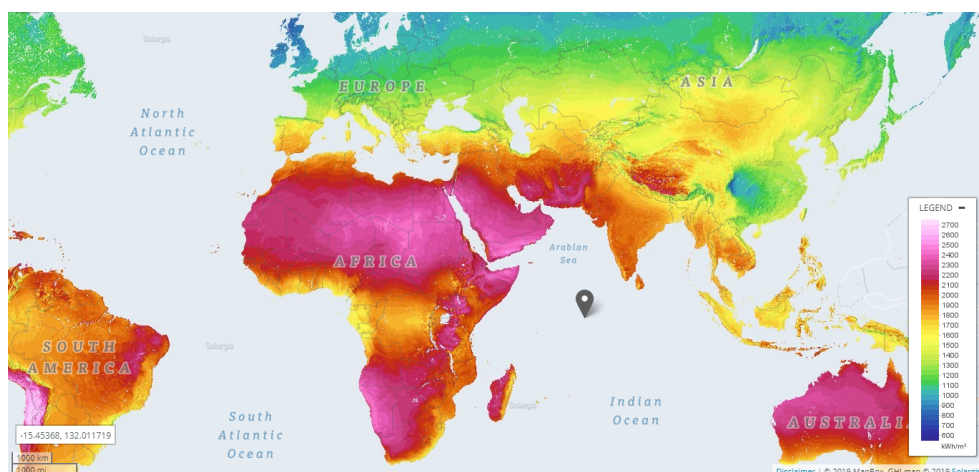


Figura 2.1: Atlas da distribuição de irradiação direta normal.

[10]

2.1.1 Componentes da radiação

As interações referentes à radiação solar são continuamente alvo de estudo e investigação, na ótica do desenvolvimento do recurso solar como fonte de energia capaz de fazer face ao uso excessivo de combustíveis fósseis e anexadas à necessidade de desenvolvimento de tecnologias de conversão de energia renovável.

No que à energia solar diz respeito, é necessário clarificar alguns conceitos que irão ser utilizados ao longo desta dissertação e cujo o seu conteúdo pode ser confundido:

- **Irradiância** — potência que incide na superfície terrestre por unidade de área (kW/m^2);
- **Irradiação** — energia que incide na superfície por unidade de área durante um determinado período de tempo. (kWh/m^2)
- **Radiação direta** — irradiação que chega à superfície terrestre proveniente de raios solares recebidos de forma direta do Sol. Incide sobre uma superfície sem mudar de direção, exceto devido à refração atmosférica;
- **Radiação difusa** — parte da radiação solar que, devido a choques sucessivos com moléculas e partículas contidas no ar, foi difundida ao atravessar a atmosfera. Radiação dependente das nuvens, apresentando grande variabilidade no tempo.
- **Radiação refletida no albedo** — radiação que chega à superfície considerada após ter sido refletida pelas superfícies envolventes. Habitualmente tem uma contribuição muito pequena e em alguns casos pode mesmo ser desprezada.
- **Radiação global** — corresponde à soma da radiação direta com a radiação difusa.

Ao penetrar na atmosfera, a radiação solar sofre fenómenos de absorção e dispersão, fenómenos que dependem da massa, espessura e densidade da massa de ar atravessada e da sua composição.^[11]

A quantidade de energia absorvida depende da massa de ar atravessada pelo raio solar, pelo que a absorção será mínima quando o sol estiver situado no ponto mais alto e máxima ao pôr do sol. Em condições de céu limpo, a quantidade de DNI que penetra na atmosfera é chamada de massa relativa do ar (AM). Quando o sol está diretamente acima de uma localização no nível do mar, $AM=1$, existe uma dependência da AM em relação ao observador, que é ilustrada na figura 2.2. Visto que o valor de $AM=1$ não pode ser alcançado em todos os locais, os espectros solares padrão de céu claro foram modelados para valores de $AM=1.5$.

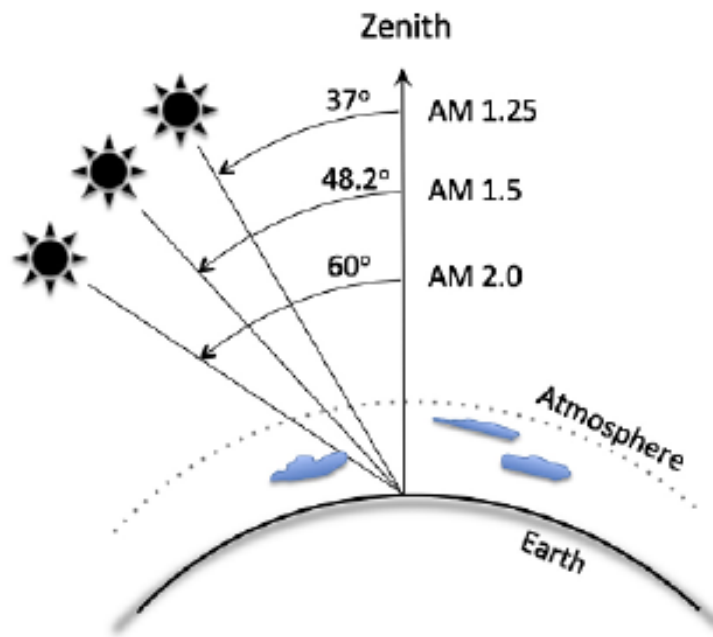


Figura 2.2: Dependência da AM em relação à posição do observador.

Por outro lado, para estudos relacionados com previsão de energia solar e avaliação de recurso, é usual utilizar os seguintes termos na divisão da radiação solar: irradiação direta normal (DNI), irradiação difusa horizontal (DHI) e irradiação global horizontal (GHI).[1]

- **DNI- Irradiação direta normal** — radiação solar direta do disco solar que incide numa superfície plana normal ao sol;
- **DHI- Irradiação difusa horizontal** — radiação solar que se dispersa por intersecção com nuvens, aerossóis e outros constituintes atmosféricos que incide numa superfície horizontal. Não inclui a DNI.
- **GHI- Irradiação global horizontal** — radiação solar hemisférica total que incide numa superfície horizontal.

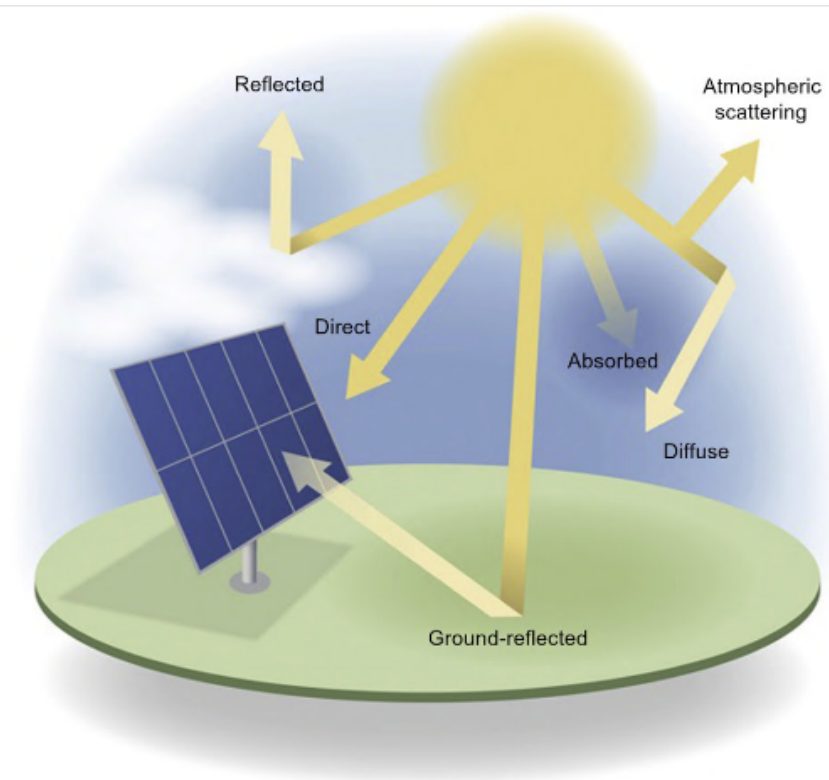


Figura 2.3: Componentes da radiação solar resultantes da interação com a atmosfera terrestre.

[1]

2.1.2 Posição do Sol: coordenadas horárias

Para o estudo do recurso solar existe a necessidade de adotar métodos que possibilitem localizar a posição solar. Um desses métodos baseia-se nas coordenadas horárias. Neste método, a posição do Sol é determinada a partir de três ângulos:

- **Declinação (δ)** — ângulo entre o plano do equador e a direção Sol-Terra. Este ângulo varia ao longo do ano, sendo nulo nos equinócios, máximo no solstício de verão e mínimo no solstício de inverno;
- **Latitude (ϕ)** — ângulo compreendido entre o paralelo do local e o equador, tem valor positivo no hemisfério norte e negativo no hemisfério sul;
- **Ângulo horário (h)** — ângulo compreendido entre o plano meridiano que passa pelo ponto considerado e o plano meridiano que passa pelo Sol. Tem a designação de ângulo horário uma vez que o seu valor depende da hora do dia.

Para o cálculo da declinação um dos métodos de cálculo utilizado é a equação dada por Lokmanhekim, uma vez que se trata duma boa aproximação:

$$\delta = 0,302 - 22,93 \cos x - 0,229 \cos x - 0,243 \cos 3x + 3,851 \sin x + 0,002 \sin 2x - 0,055 \sin 3x \quad (2.1)$$

em que δ é a declinação em radianos e:

$$x = \frac{2\pi z}{365} \quad (2.2)$$

Além desta, podemos utilizar outra equação, proposta por Cooper, na qual a aproximação não é tão confiável mas é mais sensível a desvios:

$$\delta = 23.45 \sin\left(360 \frac{284 + z}{365}\right) \quad (2.3)$$

Nesta expressão:

- δ é a declinação em graus;
- z é o número do dia do ano (1 de janeiro $z=1$ e 31 de dezembro $z=365$)

Se a velocidade da Terra fosse uniforme ao longo dum ano, todos os dias do ano teríamos 24 horas de um relógio convencional mas como isso não acontece não é possível utilizar um relógio convencional para se conhecer o tempo solar verdadeiro (TSV), expresso em horas. Para ultrapassar esta dificuldade foi definido um tempo solar médio (TSM), que é um tempo fictício, baseado na suposição de que a Terra é uma esfera que se move com uma velocidade de rotação constante ao longo do ano. A diferença entre ambos é conhecida como a equação do tempo (ET).

$$ET = TSV - TSM \quad (2.4)$$

O tempo solar médio, também conhecido como tempo civil (TC), está associado ao meridiano que passa no ponto em consideração. O tempo civil para o meridiano de Greenwich é conhecido como tempo universal (TU). Como não é possível utilizar o tempo civil de forma globalizada, uma vez que cada longitude tem o seu próprio tempo civil, a solução passou pela criação de fusos horários nos quais é conveniente que o tempo civil seja o mesmo nos pontos situados em cada uma destas zonas. Uma vez que o TSV depende da longitude e que o TC se adapta ao fuso horário, deve ter-se em conta a diferença entre a longitude do meridiano de referência e a que passa pelo ponto considerado. Além disso, em muitos países, a hora oficial (HO) não coincide com a do fuso horário, uma vez que o relógio é atrasado ou adiantado em função da época do ano. [9]

Por fim, o tempo solar verdadeiro (TSV) é calculado segundo a expressão:

$$TSV = HO - e + ET + \frac{1}{15}(\lambda_m - \lambda) \quad (2.5)$$

Nesta expressão:

- **TSV:** Tempo solar verdadeiro;

- **HO**: Hora oficial do país;
- **e**: Adianto respetivo da hora civil;
- $\frac{1}{15}(\lambda_m - \lambda)$: Fator que relaciona a longitude do meridiano de Greenwich (λ_m) e a longitude do local (λ).

A partir daqui é também possível calcular o ângulo solar. O ângulo solar varia em 15° por cada hora tendo valor nulo ao meio-dia (12 horas) de cada dia, negativo na parte da manhã e positivo na parte da tarde:

$$w = 15 \times (TSV - 12) \quad (2.6)$$

2.1.3 Posição solar: coordenadas angulares

Neste método coloca-se como referência o plano tangente à superfície terrestre no ponto considerado, em detrimento do plano equatorial, pelo que a posição solar é obtida a partir de dois ângulos:

- **Azimuth(α)** — ângulo compreendido entre a direção sul e a direção da projeção do Sol sobre o plano horizontal. A origem dos ângulos situa-se na direção Sul e consideram-se ângulos positivos os ângulos medidos na direção Oeste e negativos os ângulos medidos na direção Este;
- **Altitude(β)** — ângulo entre a posição do Sol e a sua projeção sobre o plano horizontal. Consideram-se positivos os ângulos medidos no sentido ascendente.

Por vezes também é utilizado o ângulo zenital, dado por 2.7, que corresponde ao ângulo formado entre a posição do Sol e a perpendicular ao plano horizontal no ponto considerado.

$$\psi = 90 - \beta \quad (2.7)$$

As coordenadas angulares para latitudes superiores à do trópico relacionam-se com as coordenadas horárias segundo as seguintes equações:

$$\cos \psi = \sin \beta = \sin \phi \sin \delta + \cos \phi \cos \delta \cos h \quad (2.8)$$

$$\cos \alpha = \frac{1}{\cos \beta} (\sin \phi \cos \delta \cos h - \cos \phi \sin \delta) \quad (2.9)$$

de onde se obtém:

$$\sin \alpha = \frac{\cos \delta \sin h}{\cos \beta} \quad (2.10)$$

Para um determinado ponto num dia concreto (conhece-se a sua latitude ϕ e declinação δ) poderá calcular-se a hora do pôr do sol, já que nesse instante a sua latitude é nula ($\beta=0^\circ$). Com as

expressões anteriores podemos calcular o ângulo horário de pôr de sol e o azimute correspondente:

$$\cos h_s = -\tan \phi \tan \delta \quad (2.11)$$

$$\sin \alpha_s = \cos \delta \sin h_s \quad (2.12)$$

$$(2.13)$$

Consequentemente, as horas de sol nas quais não existem nuvens podem ser calculadas segundo a expressão:

$$t_d = \frac{2}{15} \arccos(-\tan \phi \tan \delta) \quad (2.14)$$

2.1.4 Cálculo do index de claridade

A radiação solar incidente na Terra é afetada por fenómenos de dispersão, absorção ou transmissão na interseção com a atmosfera terrestre, fenómenos esses explicados pelas propriedades dos seus constituintes. Um dos intervenientes com maior influência é a nuvem, a sua importância é de tal ordem que todas as previsões de radiação solar são altamente dependentes da eficiência da previsão de nebulosidade, sendo necessária informação acerca do tipo, tamanho, movimentação e áreas de formação e dispersão.[1]

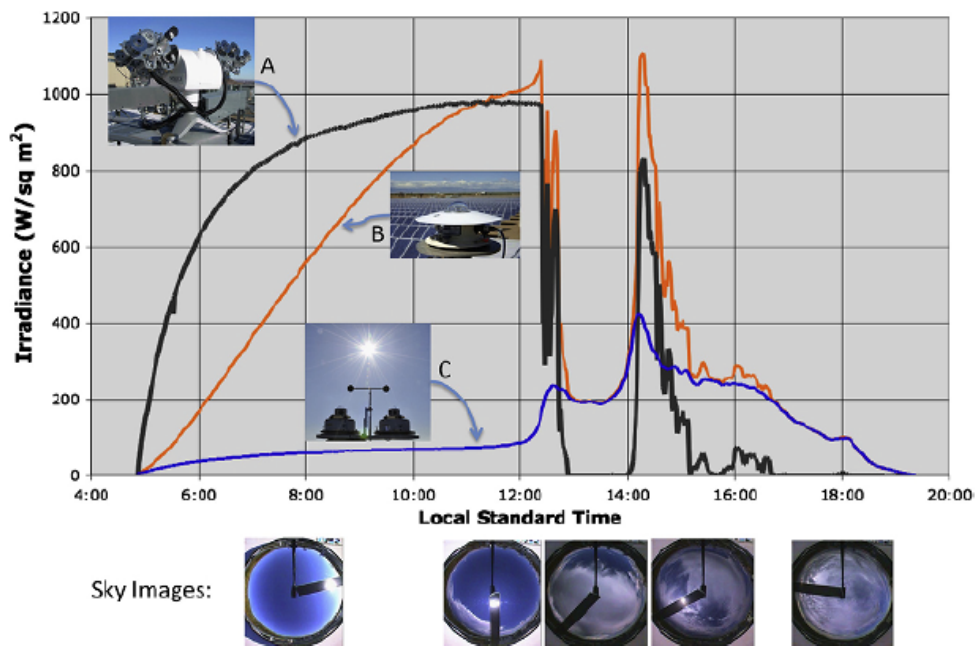


Figura 2.4: Série de tempo de componentes de irradiância solar medidas por pireliómetro (A=DNI) e piranómetros (B=GHI);C=DHI)[1]

Tendo em conta estes fenómenos, um dos índices mais utilizados na previsão solar é o chamado *índice de clareza* (Kt). O índice de clareza é definido como a razão entre a irradiância global e a

irradiância à superfície num dia de céu limpo:

$$K_t = \frac{G}{G_{clear}} \quad (2.15)$$

Nesta expressão:

- G : irradiância global;
- G_{clear} : irradiância para um dia de céu limpo.

Este índice serve como medida da transmissividade de irradiância das nuvens.[1, 12]

2.2 Estimativa do recurso solar

De uma forma geral, o processo de previsão de produção fotovoltaica é dividido em duas etapas.

A primeira etapa corresponde à previsão de radiação solar e temperatura do ar (possivelmente também outras variáveis meteorológicas, como a velocidade do vento e a humidade relativa) a partir de três famílias de métodos de previsão, escolhidos tendo em conta o *lead time* de previsão, a carga computacional e a possibilidade de adquirir medições locais:

- i) Métodos baseados em medições no solo;
- ii) Métodos baseados em dados de satélites;
- iii) Métodos baseados em modelos meteorológicos numéricos (NWP);

No nosso contexto o *lead time* corresponde à diferença entre o tempo para o qual queremos executar a previsão e o tempo no qual foram recolhidos os dados.

A segunda etapa corresponde à conversão para potência das variáveis previstas, tendo em atenção as especificações técnicas do local. A incerteza na previsão da radiação solar é quase uma ordem de grandeza maior do que a incerteza associada à conversão da radiação solar em potência fotovoltaica, pelo que a precisão na previsão da radiação solar é um fator chave neste processo. [13, 14, 15]

2.2.1 Modelos de previsão solar

Os modelos de previsão solar normalmente são utilizados tendo em conta o máximo *lead time* envolvido em cada previsão.

- **Previsões de horizonte temporal menor do que 1 hora:** Para previsões em que o horizonte temporal é sub-horário, os métodos baseados no solo são mais eficientes uma vez que nessa escala de tempo as mudanças nos padrões climáticos são relativamente pequenas e afetadas sobretudo por características locais. As previsões baseadas na combinação de medições de radiação solar e métodos estatísticos, como modelos auto-regressivos, são mais eficientes.

Neste caso é adotado um modelo simplista, denominado de "smart persistence", que admite que a posição das nuvens não muda no lead time da previsão e apenas mudanças solares determinísticas são modeladas.

Uma outra abordagem de previsão sub-horária de radiação solar baseia-se na observação do campo de nuvem a partir da localização em estudo usando um sistema de câmaras estrategicamente colocados no solo. Pela comparação de duas ou mais imagens, a trajetória e velocidade das nuvens pode ser determinada e utilizada para prever a sua localização no futuro. Esta técnica pode oferecer potencialmente uma descrição detalhada acerca da passagem das nuvens sobre a estrutura FV instalada e/ou a instalar fornecendo informações que possam resolver problemas que envolvam sombreamento de nuvens. Ainda assim, trata-se de uma abordagem relativamente recente que necessita de maior pesquisa para que possa ser usada amplamente como técnica de previsão, uma vez que existem ainda vários desafios que impedem a sua implementação. Requer um software sofisticado de armazenamento, gestão e processamento de um grande volume de dados. Normalmente o horizonte de previsão com utilização de *sky cameras* não excede os 15 minutos.

- **Previsões de horizonte temporal de algumas horas:** À medida que o *lead time* de previsão se move de sub-horas para várias horas à frente, a importância relativa dos recursos de tempo remoto prevalece sobre os recursos locais. Uma das explicações é que as nuvens longe do local em estudo irão afetar o clima local num espaço entre dezenas de minutos a poucas horas. Como consequência, as observações locais não são suficientes para explicar eventos futuros e a área de observação precisa ser expandida. Desta forma, os métodos baseados em satélites são naturalmente utilizados. A Figura 2.5 mostra que a eficiência nas previsões de métodos baseados no solo é eventualmente superada pela eficiência de previsão de métodos baseados em satélite a partir de previsões de horizonte de cerca de meia hora. Sensores implementados em satélites modernos fornecem imagens de conjuntos de nuvens ao longo de milhares de quilómetros. Estes descrevem nuvens com uma resolução espacial na ordem dos 3 km e taxas de atualizações entre 10 e 30 minutos, dependendo do tipo de satélite. Assim como a técnica de *sky cameras*, o princípio de previsão consiste na análise de similaridade de duas ou mais imagens de nuvens consecutivas. Tal como no método baseado em medições no solo, as imagens analisadas são utilizadas para determinar a velocidade e trajetória das nuvens, representadas por um campo espacial de vetores denominado de CMV (Cloud Motion Vectors). A maior limitação associada à técnica CMV (usa imagens de satélite e *sky cameras*), ocorre quando o movimento vertical das nuvens não é desprezável em relação ao deslocamento horizontal. Ao contrário das *sky cameras*, as imagens de satélite não necessitam de elevados custos de equipamento e manutenção.
- **Previsões de horizonte temporal de mais do que algumas horas:** À medida que o *lead time* de previsão aumenta, a eficiência das previsões com utilização de métodos baseados em satélites decresce. Isto acontece porque a correlação espaço-temporal entre os padrões

climáticos do presente e do futuro torna-se bastante díspar. Os métodos de previsão baseados em NWP tendem a fornecer uma eficiência de previsão mais alta do que os métodos baseados em satélites, para previsões superiores às 5 ou 6 horas seguintes. Os modelos NWP podem abranger toda a superfície terrestre — modelos globais — ou apenas áreas limitadas — modelos locais ou de mesoescala. Os modelos globais de NWP — que são executados virtualmente apenas por estações meteorológicas públicas e centros de pesquisa devido aos seus enormes esforços computacionais — fornecem cobertura mundial em detrimento da resolução espacial (atualmente na faixa aproximada de 9 a 25 km) e resolução temporal para cada hora ou a cada 3 horas. Os modelos globais NWP mais utilizados são : *Integrated Forecasting System* (IFS) do *European Centre for Medium-range Weather Forecast* (ECMWF) e o *Global Forecast System* (GFS) do *National Centers for Environmental Information* (NCEI) dos EUA.

Em contraste, os modelos de mesoescala focam-se em áreas limitadas e simulam o sistema meteorológico com maior resolução espacial e temporal. Pela redução dos requisitos computacionais por comparação com os modelos globais, estes são frequentemente utilizados por entidades privadas uma vez que se conseguem adaptar às necessidades específicas de cada cliente.

- **Previsões de longo termo:** Este tipo de previsão é requerido durante estados iniciais de projetos FV para estudos de viabilidade e bancabilidade. Resumidamente, o recurso previsto para os próximos anos e décadas é necessário para adotar um plano de negócios confiável. Curiosamente, nenhum modelo de previsão é utilizado para este tipo de aplicações. Em vez disso, dados históricos observados e dados de *typical meteorological year*, TMY, são utilizados com a suposição de que o comportamento observado no passado será similar ao futuro. Por vezes estes dados históricos são corrigidos tendo em conta erros de tendência ou desvios climáticos previstos.[16, 17]

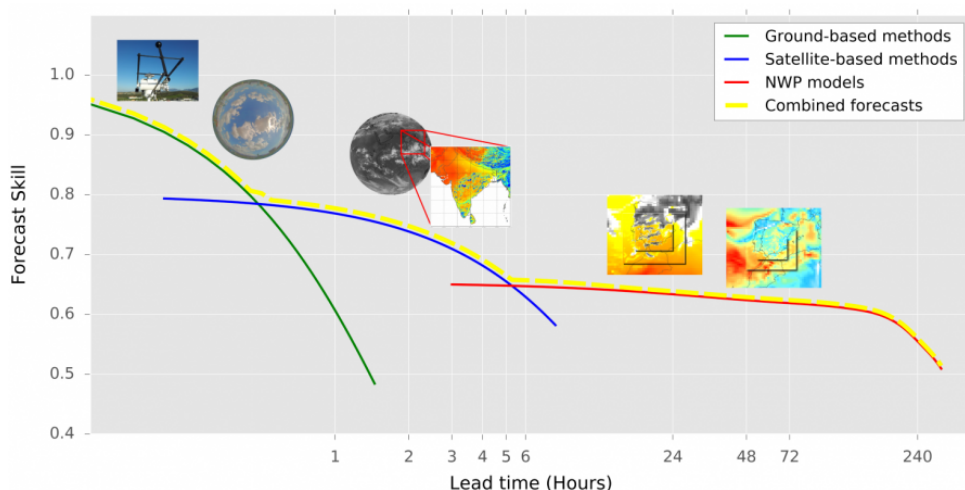


Figura 2.5: Imagem conceitual da eficiência de previsão em relação ao prazo de previsão previsto para os métodos baseados no solo, em satélite ou em NWP.

[16]

2.2.2 Previsão de radiação solar utilizando imagens obtidas por satélite (sky cameras)

Como dito anteriormente, para previsões com horizonte temporal horário, os dados obtidos por satélite são bastante mais eficientes em relação aos demais tendo em conta a sua resolução temporal e espacial. A presença de nebulosidade tem bastante impacto na radiação solar incidente na superfície, portanto, uma descrição aproximada relativamente ao deslocamento da nuvem é essencial para a previsão da radiação.

Um dos métodos mais utilizados, representado na figura 2.6, baseia-se nos vetores de deslocamento das nuvens que consiste nos seguintes passos:

1. São calculados índices de nebulosidade;
2. A partir de duas imagens consecutivas do índice de nebulosidade são calculados vetores de deslocamento que fornecem informação relativamente à direção e velocidade;
3. A previsão do comportamento das nuvens é determinado aplicando os vetores previamente calculados à imagem mais recente extrapolando assim o movimento das nuvens;
4. Aplica-se um filtro de suavização de forma a eliminar variáveis aleatórias pouco relevantes ou que não são previsíveis;
5. A radiação solar é, portanto, calculada a partir das imagens com o índice de nebulosidade previsto, utilizando o método *Heliosat*;
6. Informações da precisão para a irradiância são então fornecidas;

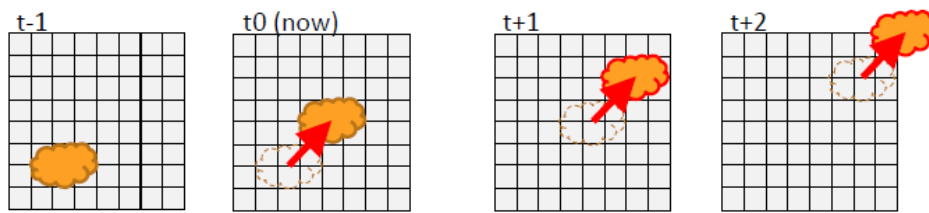


Figura 2.6: Método de vetores de deslocamento de nuvens

2.2.3 Modelos de previsão numérica de tempo (NWP)

Para previsões com horizontes temporais de alguns dias, os modelos de previsões numéricas afiguram-se como sendo a melhor alternativa. Os modelos meteorológicos como um modelo NWP, fazem a previsão das condições climatéricas utilizando as condições atuais como entradas de modelos matemáticos que simulam o comportamento da atmosfera. Estes modelos normalmente dividem-se entre modelos globais e modelos locais, sendo que os primeiros têm geralmente uma baixa resolução não permitindo um mapeamento detalhado das características em pequena escala. Portanto, são necessários modelos de mesoescala regionais juntamente com os modelos NWP para prever os efeitos locais.

2.2.3.1 Modelos de previsão numérica de tempo (NWP) globais

Os modelos globais apresentam previsões meteorológicas de maior alcance, escala mundial, com uma resolução espacial na ordem dos 200 km e fornecem essencialmente informações gerais sobre o comportamento da atmosfera. São capazes de identificar fenómenos meteorológicos de grande escala. Segue alguns exemplos:

- **GFS- *Global Forecast System*** — modelo de previsão acoplado, composto por quatro modelos separados que trabalham em conjunto para fornecer uma imagem precisa das condições meteorológicas. O GFS cobre todo o globo com uma resolução horizontal de 28 km;
- **GEFS- *Global Ensemble Forecast System*** — modelo de previsão meteorológica de cobertura global composto por 21 previsões separadas, ou membros do grupo, utilizados para quantificar a quantidade de incerteza numa previsão. O GEFS produz output quatro vezes por dia com precisões de tempo que podem chegar às 16 horas;
- **ECMWF- *European Centre for Medium-Range Weather Forecasts*** .

Para previsões a curto prazo este tipo de modelos não são utilizados, uma vez que incluem modelos matemáticos complexos, processados em supercomputadores.

2.2.3.2 Modelos de previsão numérica de tempo (NWP) locais

Os modelos locais concentram-se em grandes áreas, que podem ser continentes ou um conjunto de países e a sua resolução espacial está entre 2 km e 50 km. Têm como objetivo primordial analisar e identificar o comportamento da atmosfera sobre uma região específica.

- **NAM- *North American Mesoscale*** — modelo regional de previsão de tempo que abrange a América do Norte numa resolução horizontal até 12 km;
- **RAP- *Rapid Refresh*** — modelo regional de previsão de tempo da América do Norte com sub-redes separadas (com diferentes resoluções temporais) dentro do domínio geral da América do Norte. Gera previsões a cada hora com duração prevista de 18 horas.

Uma vez que permitem estudar uma zona específica com grande resolução espacial, este tipo de modelos são de grande importância na previsão da produção de parques fotovoltaicos e eólicos.

2.3 Redes Neurais Artificiais (ANN)

Uma das técnicas de previsão mais populares é ANN. As Redes Neurais Artificiais, uma das técnicas no âmbito da Inteligência Artificial, simulam o modo de funcionamento do cérebro humano. As ferramentas de modelização de redes neurais resolvem problemas complexos de dados não lineares, organização de dados, reconhecimento de padrões, simulação e otimização. A eficiência das previsões com o uso de modelos ANN depende dos parâmetros utilizadas como inputs, algoritmo de treino e configuração da rede (figura 2.7- estrutura básica duma rede neuronal).[2, 18, 19]

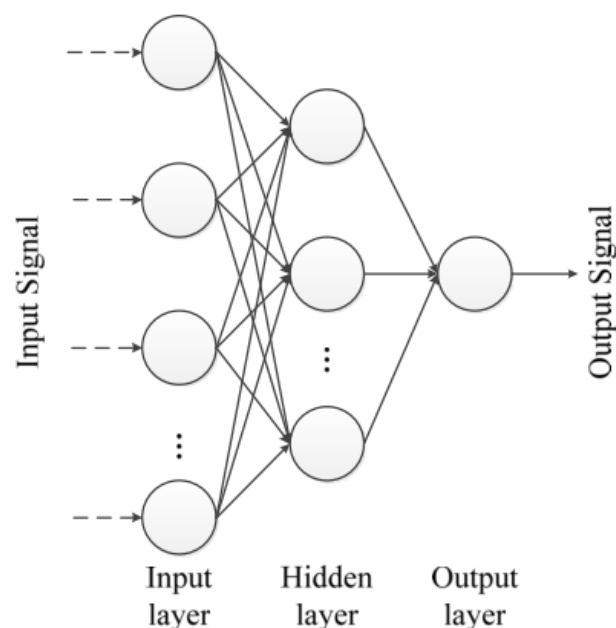


Figura 2.7: Estrutura de uma rede neuronal simples com 3 camadas [20]

A modelação da rede neuronal pode ser dividida em três fases:

- **Estrutura da rede** — tipo de rede neuronal, número de neurónios escondidos em cada camada, número de camadas, parâmetros de input e output, treino e validação dos dados seleccionados;
- **Fase de treino** — inicia quando os dados estão introduzidos na rede e o peso foi modificado até que uma condição pré-determinada seja satisfeita e finalizada;
- **Fase de validação** — quando o modelo construído pela ANN é validado a partir de dados que não foram utilizados na fase de treino.

Das estruturas presentes nas redes neuronais aquela de uso mais comum é a MLP (The Multilayer Perceptron). Esta estrutura contém uma camada de entrada, uma ou mais camadas escondidas e uma camada de saída. A camada de entrada e de saída é representada por vetores, respetivamente, x e y , e a camada escondida é caracterizada pelo número de neurónios.

2.3.1 Neurónios

Uma rede neuronal é constituída por várias unidades de processamento, denominadas de neurónios, cujo modelo é inspirado no sistema nervoso biológico.

Um neurónio índice x da camada k recebe sinais O_{pz} dos neurónios da camada anterior p e efetua as transformações apresentadas na figura 2.8, antes de produzir um sinal de saída O_{kx} que, por sua vez, irá ativar neurónios na camada seguinte:

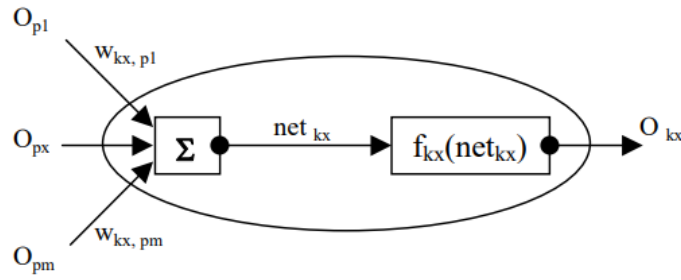


Figura 2.8: Modelo simplificado da estrutura de um neurónio [21]

$$net_{kx} = \sum_{m=1}^z w_{kx, pz} O_{pz} \quad (2.16)$$

$$O_{kx} = f_{kx}(net_{kx}) \quad (2.17)$$

De seguida, a saída do neurónio é processada pela função de ativação que normalmente é de um dos três tipos: em degrau, linear ou sigmóide.

2.3.2 Arquitectura da rede

Quanto à arquitectura das redes neuronais, estas podem ser divididas em redes unidireccionais (*feedforward*) ou recorrentes (*feedback*). As redes mais utilizadas são as redes *feedforward*. Este tipo de redes funciona de forma progressiva, em que não existe qualquer informação adicional acerca dos processos a jusante, ou seja, sem realimentação. [22]

Nas redes *feedforward*, os neurónios são organizados sob a forma de camadas, em que a informação flui a partir da camada de entrada para a camada de saída num único sentido, sem inversão do caminho. [23]

2.3.3 Processo de aprendizagem

O processo de treino ou aprendizagem de uma rede neuronal pode ser dividido em três grupos:

- **Treino supervisionado:** Neste processo é conhecido à priori qual o alvo para o qual a rede está a ser treinada. Desta forma, é construído um sistema mapeador entre os valores de entrada e saída da rede, a rede produz uma resposta baseada no conjunto de dados de entrada. Este é o tipo de treino mais utilizado nos modelos de previsão.
- **Treino não supervisionado:** Situações nas quais não são conhecidos os valores de saída. Os conjuntos de dados de entrada são mapeados num conjunto de dados de saída, a partir dum processo de classificação dos dados com base na sua semelhança. Os pesos ajustam-se por

forma a maximizar a transferência de informação da entrada para a saída, sem conhecimento acerca da qualidade da saída obtida.

- **Aprendizagem forçada:** Não são fornecidas as saídas corretas para as entradas mas são atribuídos prêmios ou castigos de acordo com a performance da saída da rede.

2.3.4 Normalização de variáveis

De forma a melhorar a eficiência do modelo e aumentar a rapidez do processo de aprendizagem, normalmente no uso de redes neurais é utilizado uma normalização das variáveis. [24] Os métodos mais utilizados de normalização são:

- **Método Min-Máx:** Quando são conhecidos os valores mínimos e máximos das variáveis de entrada, este tipo de estandardização permite normalizar os dados a partir duma transformação linear:

$$v^* = \frac{v - \min_a}{\max_a - \min_a} \times (\max_A - \min_A) + \min_A \quad (2.18)$$

Nesta expressão::

- v^* : Valor normalizado da variável de entrada;
- v : Valor original da variável de entrada;
- $\min_a$ e $\max_a$: Valor mínimo e máximo de cada variável;
- $\min_A$ e $\max_A$: Valor mínimo e máximo na escala de normalização.
- **Método Z-score:** Calcula-se a média e o desvio padrão das variáveis de entrada a normalizar. Posteriormente, as variáveis são normalizadas da seguinte forma:

$$v^* = \frac{v - \langle v \rangle}{\sigma} \quad (2.19)$$

Nesta expressão:

- v^* : Valor da variável de entrada normalizado;
- $\langle v \rangle$: Média dos valores da variável de entrada;
- σ : Desvio padrão dos valores da variável de entrada.
- **Método Decimal Scaling:** Mudança de escala com divisão por múltiplos de 10, onde j é o menor valor inteiro que satisfaz a condição:

$$\max |v^*| < 1 \quad (2.20)$$

Normalização:

$$v^* = \frac{v}{10^j} \quad (2.21)$$

2.3.5 Avaliação do desempenho dos modelos de previsão

A análise da performance dos métodos desenvolvidos pode ser avaliada segundo determinadas métricas de avaliação. Algumas pesquisas efetuadas concentraram as métricas utilizadas como base para a avaliação completa de previsão: [2, 25]

- **Pearson's correlation coefficient:** O coeficiente de correlação de Pearson mede a correlação entre dois parâmetros ou dois conjuntos de dados, seguindo a seguinte equação:

$$(p) = \frac{cov(\rho_{meas}, \rho_{pred})}{\sigma_{meas} \sigma_{pred}} \quad (2.22)$$

Nesta expressão:

- ρ_{meas} e ρ_{pred} indicam o output atual e previsto.

- **Normalized Error (nE):**

$$nE = \frac{P_{pred} - P_{meas}}{\max(P_{pred})} \quad (2.23)$$

Nesta expressão:

- P_{meas} e P_{pred} são, respetivamente, outputs medidos (meas) e previstos (pred).

- **Mean Absolute Error (MAE):** refere-se á distância média entre os valores medidos e o modelo previsto:

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_{pred} - P_{meas}| \quad (2.24)$$

- **Mean Bias Error (MBE):** métrica utilizada para calcular o bias da previsão:

$$MBE = \frac{1}{N} \sum_{i=1}^N (P_{pred} - P_{meas}) \quad (2.25)$$

- **Standard Deviation Error (SDE):**

$$SDE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{pred} - P_{meas} - MBE)^2} \quad (2.26)$$

- **Root Mean Square Error (RMSE):** Erro global medido de todo o período de previsão:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{pred} - P_{meas})^2} \quad (2.27)$$

- **Mean Absolute Percentage Error (MAPE):** Método da avaliação da capacidade que um conjunto previsto tem em se adaptar aos valores reais:

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{P_{pred} - P_{meas}}{P_{avg}} \right| \quad (2.28)$$

- **Coefficient of Determination (R^2)**: Representa a correlação entre os valores reais e previstos.

$$R^2 = 1 - \frac{Var(P_{meas} - P_{pred})}{Var(P_{pred})} \quad (2.29)$$

2.4 Métodos de MCP solar

O princípio básico da metodologia de MCP é correlacionar dados de curto termo, normalmente dados medidos de um certo local de referência, com dados de longo termo próximos de local, muitas vezes fornecidos por estações meteorológicas, para que seja estabelecida uma relação entre eles. [26]

O método mais comum de MCP é o método por regressão linear. O exemplo mais simples deste tipo de método é uma função linear em que:

$$U_{target} = U_{ref}m + b \quad (2.30)$$

Nesta expressão:

- U_{target} : Variável associada aos dados medidos;
- U_{ref} : Variável associada aos dados de Longo Termo;

Os parâmetros b e m são determinadas a partir de uma regressão linear entre as medições disponíveis do local e os dados de Longo Termo. Uma vez determinados os coeficientes, a reta pode ser aplicada ao conjunto de dados de Longo Termo para calcular os valores correlacionados correspondentes. Ainda assim, por exemplo no cálculo da correlação das velocidades do vento, a direção do vento deve ser considerada uma vez que afeta o recurso eólico de forma diferente nas medições locais e nos dados de Longo Termo. Desta forma, outros métodos não lineares baseados, por exemplo, em Redes Neurais afiguram-se como uma boa alternativa. [27, 26]

2.4.1 Métodos utilizados na empresa Megajoule

Os métodos de MCP utilizados pela empresa Megajoule, são descritos de seguida:

Variáveis utilizadas:

- **GHI**: Irradiância horizontal global (W/m^2);
- **DHI**: Irradiância horizontal difusa (W/m^2);
- **Temp**: Temperatura ambiente ($^{\circ}C$);
- **Kt**: Index de claridade: quociente entre valores de GHI e clear sky;

Dados utilizados:

- **Série de LT**: série de valores horários de GHI, DHI e Temperatura;

- **TMY**: valores horários de um ano meteorológico típico de GHI, DHI e Temperatura.
- **Dados medidos** : dados medidos 10 em 10 minutos no local em estudo de GHI e Temperatura;

Métodos utilizados para a irradiância:

- **Método KT Gamas (GHI GAMAS e DHI GAMAS)**: Para cada mês é efetuada a regressão linear entre os valores medidos e os valores da série de LT obtendo-se os coeficientes m e b. Reta, em que o y corresponde aos valores medidos e o x aos valores da série de LT. Posteriormente a equação encontrada é aplicada a cada valor horário medido correspondente ao mês.
- **Método KT Mod (GHI GAMAS e DHI GAMAS)**: Para valores de KT menores do que 0.95 são utilizadas as mesmas retas encontradas no método anterior. Para valores de KT maiores do que 0.95, é efetuado, para cada mês um rácio entre a média da irradiância dos valores medidos e da série de LT. Posteriormente, hora a hora, são aplicadas as retas correspondentes a cada mês aos valores medidos caso o kt dessa hora seja menor do que 0.95 ou o rácio caso o kt seja maior do que 0.95.

Métodos utilizados para a temperatura:

- **Método Temp**: Para cada mês é calculado um rácio entre a média da irradiância dos valores medidos e da série de LT. Posteriormente, hora a hora, são aplicadas as retas correspondentes a cada mês aos valores medidos de temperatura.

Avaliação dos métodos:

- Cálculo do *Mean Bias Error*, MBE ou BIAS, entre a média total dos valores após aplicação da correlação e a média total dos valores medidos :

$$MBE(\%) = \frac{\frac{1}{N} \sum_{i=1}^N (GHI_{pre} - GHI_{med})}{\frac{1}{N} \sum_{i=1}^N (GHI_{med})} \times 100 \quad (2.31)$$

$$MBE(\%) = \frac{\frac{1}{N} \sum_{i=1}^N (Temp_{pre} - Temp_{med})}{\frac{1}{N} \sum_{i=1}^N (Temp_{med})} \times 100 \quad (2.32)$$

Onde:

- GHI_{pre} : Registos horários de irradiância prevista;
- GHI_{med} : Registos horários de irradiância medida;
- $Temp_{pre}$: Registos horários de temperatura prevista;
- $Temp_{med}$: Registos horários de temperatura medida;
- N : Número de registos.

- Cálculo do coeficiente de determinação, R^2 , entre o vetor de valores medidos e de valores previstos:

$$R^2 = 1 - \frac{Var(GHI_{med} - GHI_{prev})}{Var(GHI_{prev})} \quad (2.33)$$

$$R^2 = 1 - \frac{Var(Temp_{med} - Temp_{prev})}{Var(Temp_{prev})} \quad (2.34)$$

- GHI_{pre} : Registos horários de irradiância prevista;
- GHI_{med} : Registos horários de irradiância medida;
- $Temp_{pre}$: Registos horários de temperatura prevista;
- $Temp_{med}$: Registos horários de temperatura medida;
- N : Número de registos;
- Var: Variância;

- Cálculo do *Mean Absolute Percentage Error*, MAPE, entre os valores medidos e os valores previstos:

$$MAPE(\%) = \frac{1}{N} \sum_{i=1}^N \left| \frac{GHI_{prev} - GHI_{med}}{GHI_{med}} \right| \times 100 \quad (2.35)$$

$$MAPE(\%) = \frac{1}{N} \sum_{i=1}^N \left| \frac{Temp_{prev} - Temp_{med}}{Temp_{med}} \right| \times 100 \quad (2.36)$$

- GHI_{pre} : Registos horários de irradiância prevista;
- GHI_{med} : Registos horários de irradiância medida;
- $Temp_{pre}$: Registos horários de temperatura prevista;
- $Temp_{med}$: Registos horários de temperatura medida;
- N : Número de registos;

Os valores de TMY só são corrigidos, tendo em conta um dos métodos, ou seja, o método em que os parâmetros de avaliação são melhores.

Capítulo 3

Processamento e Análise de dados

3.1 Introdução

Este capítulo destina-se à apresentação e caracterização dos dados associados ao caso de estudo em análise, bem como à exposição do processamento utilizado na sua sincronização. A síntese e o processamento rigoroso desta informação são pontos de partida fulcrais na construção e aplicação dos modelos de previsão.

Aqui é explicada a organização e filtragem utilizada na preparação dos dados utilizados na metodologia baseada em modelos de redes neuronais. O objetivo principal do desenvolvimento desta dissertação é a obtenção de um TMY representativo da produção típica do local do caso de estudo e compará-lo com os TMY obtidos pelos métodos de MCP já utilizados.

Neste capítulo é ainda explicada a metodologia e os processos utilizados na construção dos modelos de previsão adequados ao objetivo proposto.

Com a finalidade de encontrar uma correlação entre os dados de LT e os dados medidos, neste capítulo serão expostos os passos utilizados na construção de três modelos de Redes Neuronais (RN) para previsão de: irradiância horizontal global, index de claridade medido e temperatura ambiente. Nestes modelos serão utilizados dados de LT de anos coincidentes com a amostra de valores medidos (2014-2015) como variáveis de entrada, e dados medidos como variáveis de saída, como descrito na secção 3.2 . As próximas secções visam explicar o processo utilizado na construção das RN adequado a cada modelo, bem como a metodologia utilizada na seleção da melhor rede.

Posteriormente, as RN selecionadas foram utilizadas para a construção de dois novos TMY , que foram comparados com o TMY fornecido inicialmente pelo *software* PVGIS e com o TMY resultante da metodologia de correlação usada pela empresa Megajoule.

3.2 Caracterização dos dados utilizados

Na secção 1.4 é apresentado o tipo de dados referente ao caso de estudo exemplo utilizado na análise e aplicação da metodologia a desenvolver. Nesta secção, esses dados são caracterizados e processados:

- **Dados de Longo Termo:** registos horários de variáveis meteorológicas de longo termo para um período de 01-01-2005 a 31-12-2016;
- **Dados de TMY:** registos de médias horárias de variáveis meteorológicas referentes a um ano completo (8760 dados) obtidos a partir dos dados de longo termo do período de 01-01-2008 a 31-12-2016;
- **Dados medidos:** registos médios, máximos e mínimos medidos de de 10 em 10 minutos no local do caso de estudo em análise para o período de 01-01-2014 a 31-12-2015 e para o ano de 2017 (01-01-2017 a 31-12-2017).

3.2.1 Dados de Longo Termo

Os dados de Longo Termo utilizados foram obtidos a partir do *software* PVGIS (*Photovoltaic Geographical Information System*), que providencia dados de base históricos e dados meteorológicos de satélite de quatro fontes meteorológicas:

- PVGIS-CMSAF;
- PVGIS-SARAH;
- PVGIS-ERA5;
- PVGIS-COSMO.

Cada fonte meteorológica distingue-se pela gama de longitude e latitude de regiões que cobre e pelo período dos dados disponíveis. Neste caso , os dados utilizados são fornecidos pelo PVGIS-SARAH . O tipo de dados disponíveis no PVGIS-SARAH são obtidos pelas observações dos canais visíveis a bordo dos satélites geoestacionários *Meteosat*. São dados válidos para os anos de 2005 a 2016 e cobrem regiões dentro da gama +/- 65° de longitude e +/- 65° de latitude. Os dados resultantes podem ser médias mensais, diárias e horárias para uma rede regular de latitude/longitude com resolução espacial de 0.05 x 0.05.[28]

No nosso caso, os dados adquiridos são dados horários que englobam as seguintes variáveis:

- **LT₁:** Irradiância horizontal global (W/m^2);
- **LT₂:** Elevação solar (°C);
- **LT₃:** Temperatura ambiente (°C);
- **LT₄:** Velocidade do vento a 10 m (m/s).

A tabela 3.1 apresenta a gama de valores associada a cada uma das variáveis presentes nos dados de LT:

Tabela 3.1: Descrição de valores dos dados de Longo Termo

Designação	Gama de Valores
LT₁	[0 - 1012,6] (W/m ²)
LT₂	[-6,26 - 73,75] (°C)
LT₃	[7,57 - 25,02] (°C)
LT₄	[0,06 - 23,89] (m/s)

3.2.2 Dados de TMY

Os dados de TMY são fornecidos diretamente pelo *software* PVGIS, baseados nos dados da série de Longo Termo. Para cada mês do ano, são selecionados os dados de Longo Termo do ano considerado mais "típico" para cada mês. Por exemplo, janeiro pode corresponder aos dados de Longo Termo de 2007, fevereiro de 2012 e por aí fora. A seleção de cada mês do TMY é feita a partir do método utilizado pelo internacional Standard ISO 15927-4, baseado na temperatura do ar, humidade relativa e irradiância horizontal global.[15]

Para o caso em estudo, a compilação de anos representativos de mês típico do TMY inicial é a seguinte:

Tabela 3.2: Anos escolhidos como "mais típicos" para cada mês do ano

Mês	Ano Escolhido
janeiro	2008
fevereiro	2010
março	2013
abril	2010
maio	2016
junho	2012
julho	2011
agosto	2016
setembro	2010
outubro	2015
novembro	2012
dezembro	2016

As variáveis de TMY fornecidas foram as seguintes:

- **TMY₁**: Temperatura Ambiente (°C);
- **TMY₂**: Humidade Relativa (%);
- **TMY₃**: Irradiância Horizontal Global (W/m²);
- **TMY₄**: Irradiância Direta Normal (W/m²);

- **TMY₅**: Irradiância Difusa Horizontal (W/m^2);
- **TMY₆**: Radiação Infravermelha (W/m^2);
- **TMY₇**: Velocidade do vento (m/s);
- **TMY₈**: Direção do vento ($^\circ\text{C}$);
- **TMY₉**: Pressão do ar (Pa);

A tabela 3.3 representa a gama de valores de cada uma das variáveis:

Tabela 3.3: Descrição de valores dos dados de TMY

Designação	Gama de Valores
TMY₁	[8,22 - 24,38] ($^\circ\text{C}$)
TMY₂	[54,3 - 100] (%)
TMY₃	[0 - 985] (W/m^2)
TMY₄	[0 - 989,11] (W/m^2)
TMY₅	[0 - 469] (W/m^2)
TMY₆	[278,7 - 427,05] (W/m^2)
TMY₇	[0,57 - 23,89] (m/s)
TMY₈	[0 - 360] ($^\circ\text{C}$)
TMY₉	[97870 - 103658] (Pa)

O comportamento dos valores de temperatura ambiente e irradiância horizontal global são ilustrados nas figuras seguintes:

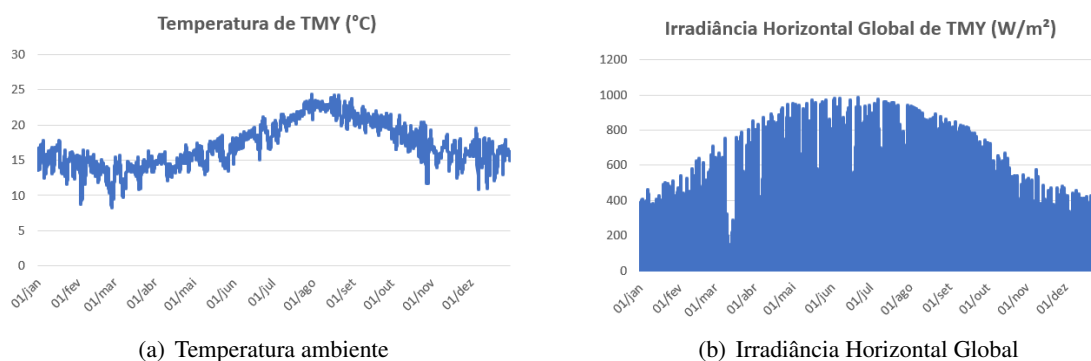


Figura 3.1: Comportamento das variáveis temperatura e irradiância nos registos de TMY

3.2.3 Dados medidos

Neste caso de estudo são fornecidos dados de irradiância horizontal global e de temperatura ambiente medidos, respetivamente, por piranómetros instalados horizontalmente e sensores de temperatura. As variáveis fornecidas foram:

- M_1 : Irradiância média horizontal global (W/m^2);
- M_2 : Irradiância máxima (W/m^2);
- M_3 : Irradiância mínima (W/m^2);
- M_4 : Temperatura média ($^{\circ}C$);
- M_5 : Temperatura máxima ($^{\circ}C$);
- M_6 : Temperatura mínima ($^{\circ}C$);

Toda a gama das variáveis dos dados medidos está representada na tabela 3.4.

Tabela 3.4: Descrição de valores medidos

Designação	Gama de Valores
M_1	[0 - 1136] (W/m^2)
M_2	[0 - 1374] (W/m^2)
M_3	[0 - 1047] (W/m^2)
M_4	[6,96 - 27,24] ($^{\circ}C$)
M_5	[7,47 - 27,71] ($^{\circ}C$)
M_6	[6,37 - 26,92] ($^{\circ}C$)

3.2.4 Dados de céu limpo (Clear sky)

A criação de modelos de clear sky permite estimar a radiação solar terrestre no caso de não haver nebulosidade, como descrito na secção 2.1.4. Desta forma, utilizando o o *software* Resun foi obtido o clear sky representativo do caso de estudo, a partir da elevação solar e altitude do local. Posteriormente este modelo de clear sky foi comparado com os dados medidos por forma a invalidar registos com erros. O clear sky obtido corresponde a valores de 10 em 10 minutos de irradiância horizontal global representativo de um ano completo.

3.3 Sincronização de dados

Pretende-se nesta fase adoptar a mesma referência de tempo para os dados apresentados anteriormente. Tendo em conta o objetivo de encontrar uma correlação entre os dados de LT e os dados Medidos, apenas estes dados são utilizados. Os passos principais para a sincronização dos dados foram os seguintes:

- Utilização dos valores de LT e Medidos dos anos de 2014 e 2015;
- Médias dos dados Medidos de 10 em 10 minutos para um regime horário de modo a sincronizar os dados com os dados de LT que já se encontram apresentados hora a hora;

- Médias dos valores de Clear Sky de 10 em 10 minutos para um regime horário de um ano completo. Aplicação do Clear Sky anual resultante, aos dois anos a sincronizar (2014 e 2015);
- Exclusão de horas com valores de erro. Valores das horas 11, 12, 13 e 14 do dia 30-01-2014 são excluídos na sincronização por apresentarem erros na medição (-999);

Os gráficos seguintes mostram o comportamento do clear sky, irradiância de Longo Termo e irradiância medida, após a sincronização, para um dia de inverno e um dia de verão.

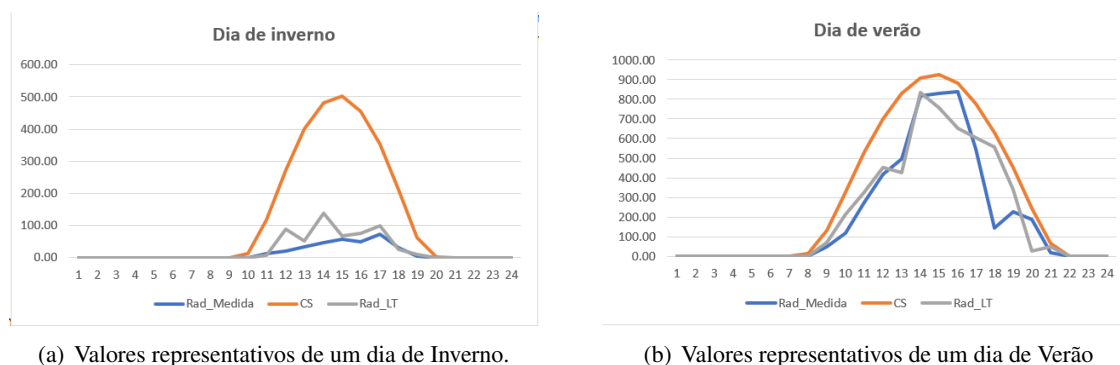


Figura 3.2: Clear sky, irradiância de longo termo e irradiância medida num dia de verão e num dia de inverno.

3.3.1 Filtragem e processamento dos dados de treino e de teste

Após efetuar o sincronismo dos dados, estes foram divididos em dois subconjuntos: subconjunto de teste e subconjunto de treino. Uma vez que a metodologia utilizada para previsão dos modelos irá ser baseada em redes neuronais, descrição completa na secção 3.5, decidiu-se dividir os dados sincronizados nestes dois subconjuntos de forma a manter um subconjunto de teste fixo no qual seja possível avaliar e comparar os desempenhos da rede.

Um dos maiores problemas durante o treino das redes neuronais é a sobreajustação. Ou seja, a rede ajusta-se de forma a que o erro do conjunto de treino seja baixo mas quando novos dados são apresentados à rede, esta apresenta erros elevados. A rede memoriza os exemplos de treino e não aprende a generalizar para novas situações.[29]

Com vista a evitar este tipo de problemas, preservou-se à partida um conjunto de testes de referência independente do treino da rede, no qual a avaliação das redes é efetuada. Na filtragem e processamento dos dados de treino e teste, todos os dados sincronizados são divididos em dois subconjuntos fixos: conjunto de treino e conjunto de teste.

Os subconjuntos são separados atendendo à semana na qual os dados estão representados, ou seja, o conjunto de teste é construído retirando uma semana de 5 em 5 semanas dos dados sincronizados, sendo os restantes valores utilizados no conjunto de treino. Os subconjuntos são então construídos da seguinte forma:

- **Conjunto de teste** — Semanas do ano: 2, 7, 12, 17, 22, 27, 32, 37, 42, 47, 52;
- **Conjunto de treino** — Semanas do ano: restantes semanas de 1 a 53.

Para modelos de treino de RN em que o *target* está associado a dados de irradiância, decidiu-se eliminar dos subconjuntos anteriormente mencionados os valores horários nulos, normalmente associados a valores noturnos em que a irradiância é inexistente. Ao mantermos estes valores estaríamos a treinar uma rede com uma grande quantidade de valores nulos na saída, sabendo à partida que estes, em circunstâncias normais, são nulos. Assim sendo, para estes casos, o número de registos horários do subconjunto de treino é de 6754 e do subconjunto de teste de 1771, atendendo à divisão explicada anteriormente, perfazendo o total de dados horários de dois anos. A figura seguinte representa um diagrama simplificado deste processo:

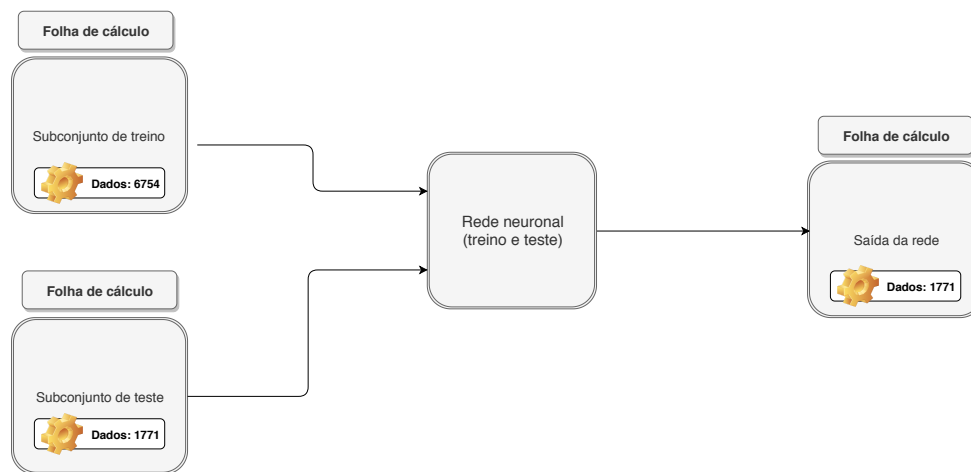


Figura 3.3: Dados utilizados nas redes neuronais para modelos de irradiância

Nos modelos associados à temperatura, como não existem valores de saída nulos (gamas de temperatura dos valores medidos de $[6,96 \text{ a } 27,24]^{\circ}\text{C}$), foram utilizados todos os dados disponíveis na divisão do subconjunto de teste e de treino. O subconjunto de treino tem 13820 registos horários e o subconjunto de teste 3698. À semelhança do referido para os modelos de previsão de irradiância é apresentado de seguida um diagrama referente ao modelo de RN para previsão de temperatura:

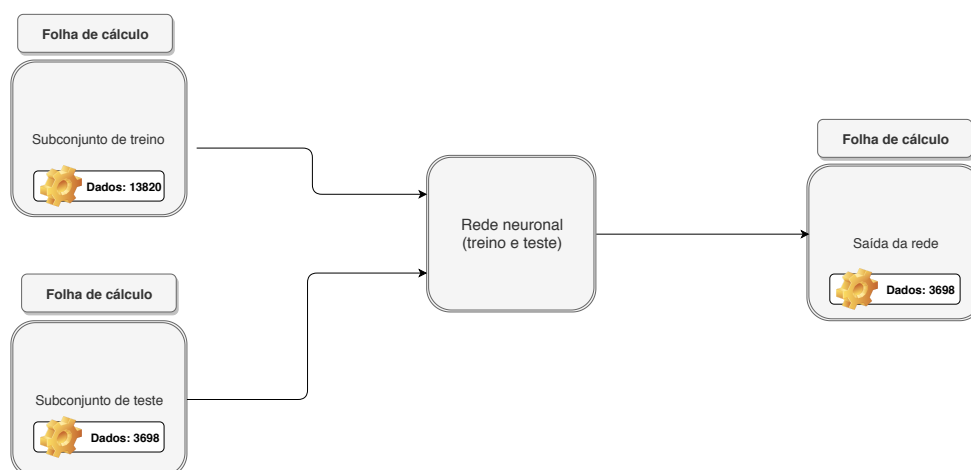


Figura 3.4: Dados utilizados nas redes neurais para modelos de temperatura

3.4 Software Utilizado

No desenvolvimento da metodologia de previsão foram utilizadas principalmente duas plataformas: o *Microsoft Excel* e o *Matlab*, este último manuseado como ambiente de programação, auxiliado pela ferramenta *nftool* utilizada para o treino, teste e validação das RN. O *Microsoft Excel* permitiu sintetizar e manipular os dados, bem como importar e exportar informação para/via *Matlab*.

3.4.1 Nftool

Por defeito, as redes neurais criadas a partir da ferramenta *nftool* utilizam redes *feedforward* com função de ativação do tipo tan-sigmóide na camada oculta da rede e função de ativação linear na camada de saída, como representando na figura 3.5.[29]

As redes *feedforward* caracterizam-se por funcionar de forma progressiva, sem realimentação, tratando-se de um aproximador universal, no qual é sempre possível encontrar uma arquitetura e um conjunto de pesos que produzam uma rede para aproximação de uma dada função.[21] Na camada oculta referida anteriormente são atribuídos por defeito 10 neurónios, valor que pode ser alterado de acordo com a performance de treino da rede. O aumento do número de neurónios na camada escondida ainda que permita a resolução de problemas mais complexos, requer um maior esforço de computação e poderá levar ao *overfitting* da rede, uma vez que esta terá tendência a ajustar-se aos dados. Na camada de saída, existe apenas um neurónio uma vez que existe apenas um alvo associado a cada vetor de entrada.[29]

Por defeito, os vetores de entrada e de saída da rede são divididos de forma aleatória em que 70 % dos dados são utilizados no conjunto de treino, 15 % no conjunto de teste e 15 % no conjunto de validação. O conjunto de treino é utilizado para calcular o gradiente e atualizar os pesos e biases da rede. O conjunto de validação permite monitorizar a performance da rede durante o processo de treino, o erro de validação normalmente decresce na fase inicial do treino, aumentando à medida

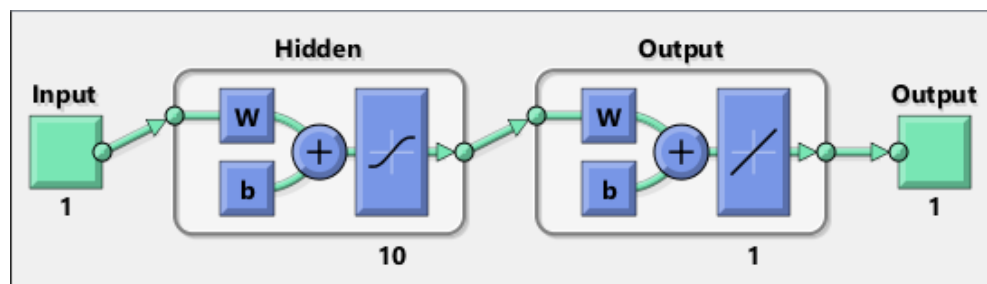


Figura 3.5: Rede de treino representativa das entradas da rede, camada escondida, camada de saída e saída da rede.

que a rede começa a ajustar os dados. O conjunto de teste não é utilizado durante o processo de treino mas é utilizado posteriormente na comparação de diferentes modelos.

Nesta ferramenta, o treino da rede é feito utilizando o algoritmo *Levenberg-Marquardt (LM)* [30], uma vez que é frequentemente o algoritmo de retropropagação mais rápido e, portanto, recomendado como o algoritmo supervisionado a utilizar, embora exija um esforço de memória superior a outros algoritmos. Este algoritmo é uma aproximação do método de Newton e é utilizado por definição neste software.

3.5 Construção da rede neuronal

A construção de uma rede neuronal deve ser elaborada tendo em conta uma série de passos importantes apresentados de seguida:

- Identificar modelos: variáveis a prever e potenciais entradas;
- Escolha de variáveis de entrada;
- Treino e teste de redes com variáveis escolhidas;
- Definição da estrutura e topologia da rede variando o número de unidades da camada escondida e entradas da rede;
- Avaliação das redes;

3.5.1 Identificação de modelos de RN: variáveis a prever e potenciais entradas

3.5.1.1 Modelos

Tendo em conta os dados de LT fornecidos, apresentados na secção 3.2.1 e os dados medidos apresentados na secção 3.2.3, e após a sincronização explicada na secção 3.3, foram propostos 3 modelos de RN para previsão de 3 saídas diferentes: Irradiância Horizontal Global Medida (Rad_M), Temperatura Ambiente Medida (Temp_M) e Index de Claridade Medido (Kt_M). Assim sendo, tendo em conta os pressupostos anteriores as possíveis variáveis utilizadas na construção são:

X1: Hora do dia	Y1: Irradiância Medida
X2: Semana do ano	Y2: Temperatura Medida
X3: Mês	Y3: Index de claridade Medido
X4: Irradiância de Longo Termo	
X5: Temperatura de Longo Termo	
X6: Index de claridade de Longo Termo	
X7: Elevação solar	

A figura 3.6 apresenta de forma ilustrativa a estrutura destes modelos. Ainda neste subcapítulo serão avaliadas as potenciais entradas dos modelos apresentados.

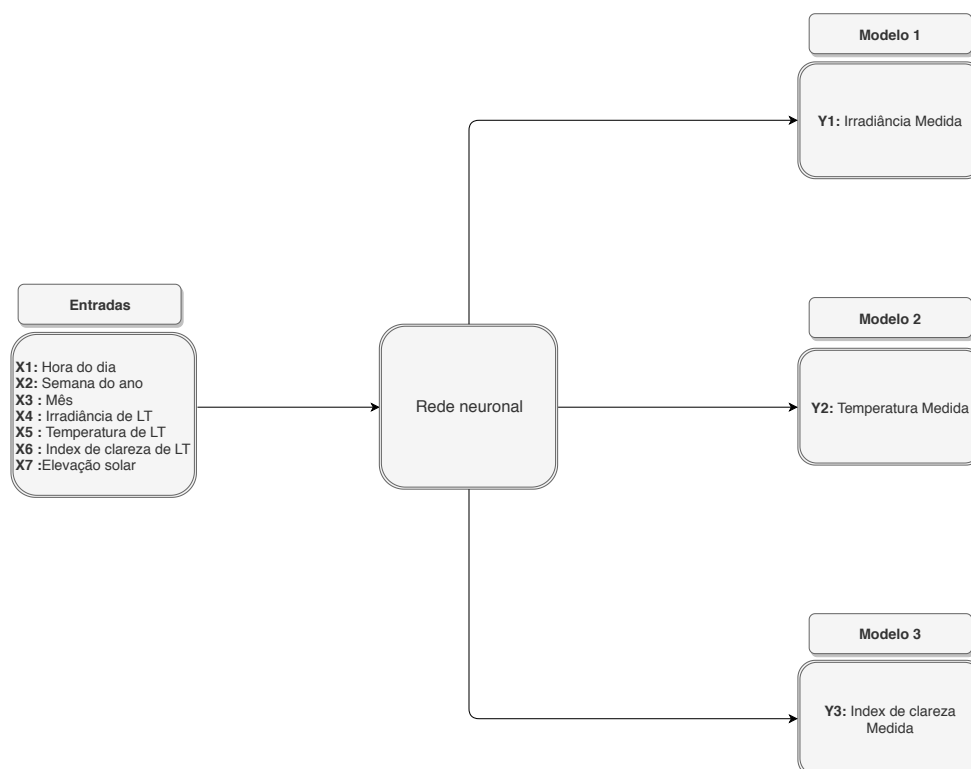


Figura 3.6: Treino dos 3 modelos: irradiância, temperatura e index de claridade

3.5.1.2 Identificação de potenciais entradas

Por forma a evitar colinearidade e informação redundante, uma das premissas na escolha do vetor de entradas das RN deve ser a garantia de independência entre essas mesmas variáveis. Posto isto, a tabela 3.5 representa o coeficiente de correlação de Pearson verificado entre elas:

Tabela 3.5: Coeficientes de correlação de Pearson entre as possíveis variáveis de entrada.

	X1	X2	X3	X4	X5	X6	X7
X1	1.000	0.000	0.000	0.230	0.023	0.193	0.264
X2	0.000	1.000	0.996	-0.029	0.364	0.027	-0.042
X3	0.000	0.996	1.000	-0.093	0.308	0.040	-0.117
X4	0.230	-0.029	-0.093	1.000	0.168	0.018	0.910
X5	0.023	0.364	0.308	0.168	1.000	0.029	0.169
X6	0.193	0.027	0.040	0.018	0.029	1.000	-0.196
X7	0.264	-0.042	-0.117	0.910	0.169	-0.196	1.000

Como se pode verificar na tabela 3.5, as variáveis X3 e X2 que correspondem, respetivamente, às entradas : mês e semana do ano, e as variáveis X4 e X7, que correspondem à irradiância horizontal global de LT e à elevação solar de LT, são variáveis com grande coeficiente de correlação. Consequentemente, uma delas foi eliminada do vetor de entrada das RN de acordo com a sua significância em relação à variável de saída de cada um dos 3 modelos propostos.

Para cada modelo, as entradas a comparar foram utilizadas como entradas de uma rede neuronal simplificada, com apenas uma entrada e uma saída, de forma a encontrar o coeficiente de correlação entre elas. A tabela 3.6 representa os resultados obtidos do coeficiente de correlação (R) das comparações mencionadas:

Tabela 3.6: Coeficientes de correlação entre as possíveis variáveis de entrada e as variáveis de saída dos 3 modelos (R)

	Y1	Y2	Y3
X2	0.284	0.842	0.052
X3	0.273	0.824	0.047
X4	0.955	0.391	0.333
X7	0.904	0.384	0.327

Pela tabela 3.6 é possível extrair que as variáveis semana do ano (X2) e irradiância de Longo Termo (X4) têm coeficientes de correlação superiores aos seus pares por comparação com as 3 saídas associadas a cada modelo da RN. Deste modo, as variáveis que serão eliminadas do vetor de entrada dos modelos de RN são a variável mês (X3) e a variável elevação solar (X7).

3.5.2 Normalização

Como referido na secção 2.3.4 a normalização das variáveis de entrada das redes neuronais permite a equalização das condições de treino e o combate do efeito de escala e de tendência. A normalização foi efetuada de acordo com a natureza das entradas:

- As entradas cronológicas, devido ao seu carácter cíclico, foram transformadas em coseno e seno, de forma a facilitar a aprendizagem. As entradas X1 e X2 tomam então valores normalizados X'_{1s} , X'_{1c} , X'_{2s} e X'_{2c} .

$$X'_{1c} = \cos\left(\frac{2\pi X_1}{24}\right) \quad (3.1)$$

$$X'_{1s} = \sin\left(\frac{2\pi X_1}{24}\right) \quad (3.2)$$

$$X'_{2c} = \cos\left(\frac{2\pi X_2}{53}\right) \quad (3.3)$$

$$X'_{2s} = \sin\left(\frac{2\pi X_2}{53}\right) \quad (3.4)$$

- As variáveis meteorológicas são normalizadas pelo método *Min-Máx*, visto tratar-se do melhor método de estandardização de variáveis quando são conhecidos os seus valores máximos e mínimos. Por conseguinte, as variáveis de entrada X4 e X5 são normalizadas para os valores de X4' e X5', de acordo com as equações:

$$X'_4 = \frac{X_4 - \min_a}{\max_a - \min_a} \times (\max_A - \min_A) + \min_A \quad (3.5)$$

$$X'_5 = \frac{X_5 - \min_a}{\max_a - \min_a} \times (\max_A - \min_A) + \min_A \quad (3.6)$$

Nesta expressão:

- $\min_a$ e $\max_a$: Valor mínimo e máximo de cada variável;
- $\min_A$ e $\max_A$: Valor mínimo e máximo na escala de normalização.
- Para além destas normalizações, a entrada X6 correspondente ao index de claridade, é obtida tendo em conta o quociente entre a irradiância de longo termo e o clear sky correspondente, como descrito na secção 2.1.4, por consequência, pode-se dizer que esta entrada é também normalizada, para X6', tendo em conta a seguinte equação:

$$X'_6 = \frac{rad_h^{LT} + 100}{cs_h + 100} \quad (3.7)$$

Nesta expressão:

- rad_h^{LT} : Registo de Irradiância horária de Longo termo.
- cs_h : Registo de clear sky horário.

De realçar ainda que para o modelo 3 , o output é também ele normalizado de forma similar à equação anterior, em que:

$$Y'_3 = \frac{rad_h^M + 100}{cs_h + 100} \quad (3.8)$$

Em que:

- rad_h^M : Registo de Irradiância horária medida;
- cs_h : Registo de clear sky horário.

Cálculo do index de claridade horário dos valores Medidos e dos valores de LT:

$$kt_h^M = \frac{rad_h^M + 100}{cs_h + 100} \quad (3.9)$$

$$kt_h^{LT} = \frac{rad_h^{LT} + 100}{cs_h + 100} \quad (3.10)$$

De modo a não existir qualquer erro de divisão por zero, nas horas em que o clear sky é nulo, foi adicionada uma constante de igual valor ao divisor e ao dividendo na fórmula do cálculo do index de claridade ,não implicando nenhuma alteração significativa ao valor final.

3.5.3 Avaliação das variáveis de entrada da rede

Por forma a avaliar as variáveis de entrada significativas a introduzir na RN, foi utilizado o *stepwise method*, que consiste na adição e/ou remoção passo a passo de uma das possíveis variáveis de entrada da rede, para posterior análise do efeito ocorrido na saída. [31]

As duas abordagens mais usuais deste método são a *Forward stepwise* e a *Backward stepwise*:

- **Forward stepwise**: são gerados n modelos, em que n corresponde ao número de variáveis de entrada em estudo. Cada um dos modelos inclui apenas uma dessas variáveis, no qual é analisado o erro *MSE* da rede. Posteriormente são criados n-1 modelos, de duas variáveis, combinando a variável de entrada cujo erro foi menor no processo anterior, com as restantes variáveis. Este processo é repetido, aumentando o número de variáveis de entrada, até que todas as variáveis de entrada sejam incluídas. A ordem de inclusão das variáveis de entrada corresponde à ordem de importância da sua contribuição.
- **Backward stepwise**: são criados n modelos, usando em cada um deles apenas n-1 variáveis disponíveis. A variável em falta no modelo com maior erro corresponde à variável mais importante. Posteriormente são criados n-1 modelos, cada um com n-2 variáveis, uma vez que para além da variável já eliminada são retiradas uma das outras variáveis em cada modelo. Este processo repete-se diminuindo o número de variáveis de entrada em cada modelo, até

que todas as n variáveis sejam eliminadas. A ordem de eliminação das variáveis corresponde à ordem de importância da sua contribuição na rede.

Na avaliação da entradas das redes foram utilizados os dois métodos por forma a perceber que combinação de entradas resultaria num melhor desempenho.

3.5.4 Algoritmo desenvolvido para teste e treino de rede

Neste subcapítulo é descrito o funcionamento do algoritmo desenvolvido para identificação das melhores redes, tendo em conta a variação das variáveis de entrada e o número de neurónios da camada escondida. O processo iterativo utilizado neste algoritmo é descrito de seguida:

- **Passo 1:** Leitura dos dados a partir da folha do *Microsoft Excel* considerando as restrições analisadas no capítulo 3.3.1 ;
- **Passo 2:** Escolha do número de neurónios da camada escondida. Variável *hiddenLayerSize* alterada por um processo de "tentativa/erro";
- **Passo 3:** Treino e teste de 10 redes independentes entre si;
- **Passo 4:** Cálculo de um *output* final, resultante da média aritmética dos outputs de cada uma das redes mencionadas anteriormente;
- **Passo 5:** Ciclo com repetição dos 2 últimos passos 10 vezes, obtendo-se uma matriz com 10 *outputs* de médias aritméticas de 10 redes;
- **Passo 6:** Cálculo do MAPE (*Mean Absolute Percentage Error*) das 10 repetições e do desvio padrão dessas 10 repetições;
- **Passo 7:** Exportação para folha do *Microsoft Excel* dos valores de MAPE de cada uma das repetições, média dos valores de MAPE e desvio padrão.

O facto de no **passo 4** serem utilizadas 10 redes para determinar o *output* final, permite que as comparações efetuadas para o método explicado na secção 3.5.3 sejam mais fidedignas pelo facto de estarmos a comparar médias de 10 redes com as mesmas características (variáveis de entrada e número de neurónios da camada escondida). Além disso, o MAPE enunciado no **passo 6** é calculado para um subconjunto de teste fixo que nada tem a ver com o treino da rede, como explicado na secção 3.3.1. Este MAPE é então calculado utilizando a expressão :

$$MAPE = \frac{\frac{1}{n_h} \sum_h |O_h - O_h^{RN}|}{\frac{1}{n_h} \sum_h O_h} \quad (3.11)$$

Nesta expressão:

- O_h : Valores horários reais do subconjunto de teste;

- O_h^{RN} : Valores horários de saída das redes neuronais;
- n_h - Total de registos do subconjunto de teste e da saída das redes neuronais.

Posteriormente, os dados exportados para a folha do *Microsoft Excel*, relatados no **passo 7** são utilizados para comparar a os desempenhos das diferentes redes através dum processo explicado na secção 3.5.5. É apresentado de seguida um exemplo dos valores finais:

Tabela 3.7: Exemplo de dados exportados após terminar o algoritmo

	Rede 6 Nº de Inputs:4 (-X5) Output: Y1 Neurónios:10
Mape 1	20,36 %
Mape 2	20,33 %
Mape 3	20,45 %
Mape 4	20,37 %
Mape 5	20,47 %
Mape 6	20,41 %
Mape 7	20,36 %
Mape 8	20,42 %
Mape 9	20,40 %
Mape 10	20,42 %
Mape_médio	20,40 %
Desv_padrão	0,04 %

Em relação a esta tabela:

- **Rede 6** representa o número de redes já testadas para esse modelo;
- **Nº de Inputs:4 (-X5)** representa o número de inputs utilizados nesta RN (4) e os inputs que foram descartados (-X5), neste caso a Temperatura de Longo Termo;
- **Output:Y1** representa o modelo e a saída para o qual a rede foi treinada, ou seja, Modelo 1 saída Y1, referente à Irradiância Horizontal Global Medida;
- **Neurónios:10** corresponde ao número de neurónios utilizados na camada escondida nesta arquitetura da rede;
- **MAPE 1 (...) MAPE 10**: valores dos MAPE's associados ao descrito nos **passos 3 a 6** do algoritmo;
- **MAPE médio**: representa a média dos MAPE's mencionados no item anterior;
- **Desv padrão** corresponde ao desvio padrão dos MAPE's já citados.

Posto isto, é representado na figura 3.7 um fluxograma representativo do algoritmo desenvolvido.

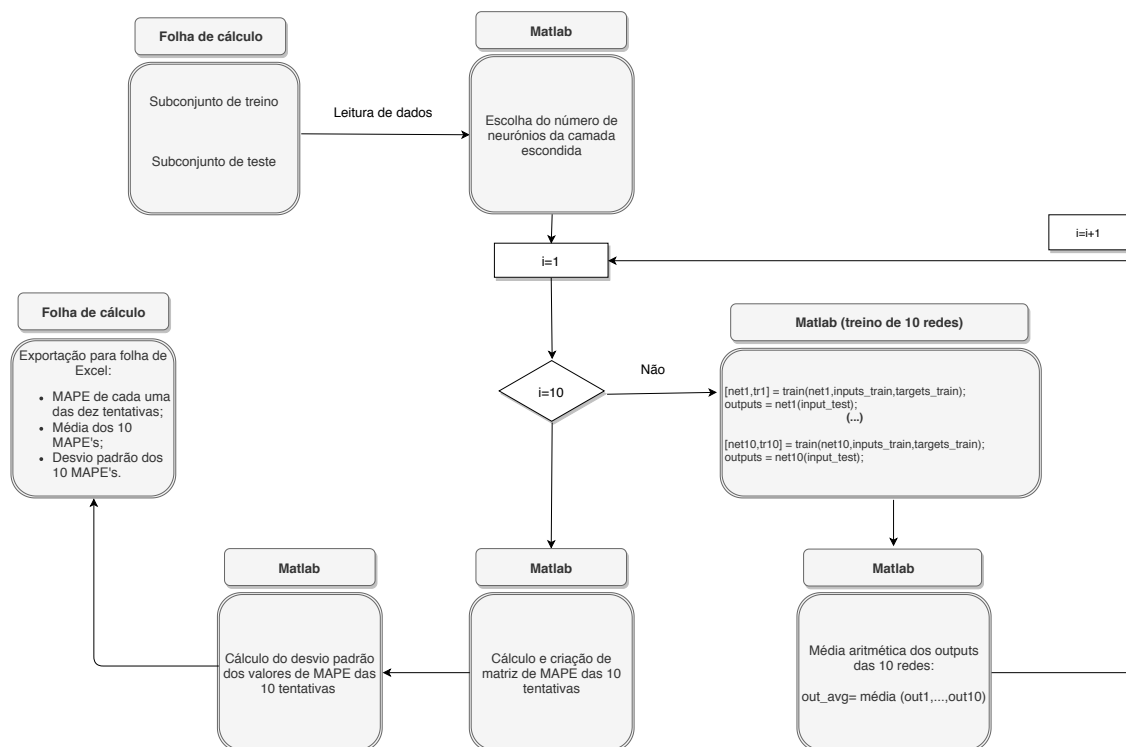


Figura 3.7: Fluxograma do algoritmo utilizado

3.5.5 Avaliação de redes

No **passo 6** da descrição do algoritmo da secção 3.5.4, referiu-se que para cada rede exportada do *software Matlab* estaria representada a média de Mape's e o desvio padrão. Posto isto, é possível comparar essas redes tendo em conta métodos estatísticos. Se, hipoteticamente, considerarmos a rede A e a rede B, é então possível determinar qual a probabilidade de "A ser melhor do que B". Esta avaliação é efetuada utilizando a função de distribuição acumulada dada pela expressão 3.12:

$$F(x) = P(X \leq x) \quad (3.12)$$

A função F é igual á probabilidade de uma variável aleatória X assumir um valor igual ou inferior a determinado x. Uma vez que a variável aleatória X é uma variável contínua, a função de distribuição acumulada pode ser representada por:

$$F(x) = \int_{-\infty}^x f(x_i) dx \quad (3.13)$$

Nesta expressão, $f(x_i)$ representa a função de densidade de probabilidade.

A intersecção no eixo dos xx's das retas de frequência acumulada das duas redes, em que:

$$1 - F_B(x) = F_A(x) \quad (3.14)$$

Permite encontrar o valor de x para o qual a função de frequência acumulada da rede A intersestará a probabilidade da "rede A ser melhor do que a rede B".

De seguida, é apresentado um exemplo representativo desta metodologia. A tabela 3.8 representa valores de duas redes:

Tabela 3.8: Exemplo da comparação de duas redes

	Rede 1 Inputs:4 (-semana) Output: Y3 Neurónios:10	Rede 2 Inputs:5 Output: Y3 Neurónios:5
Mape 1	20.28%	20.33%
Mape 2	20.29%	20.32%
Mape 3	20.31%	20.31%
Mape 4	20.23%	20.32%
Mape 5	20.25%	20.45%
Mape 6	20.24%	20.42%
Mape 7	20.32%	20.31%
Mape 8	20.25%	20.22%
Mape 9	20.22%	20.32%
Mape 10	20.28%	20.28%
Mape_médio	20.27%	20.33%
Desv_padrão	0.04%	0.06%

Pela aplicação da equação 3.14, o valor de x encontrado foi de 20.29%:

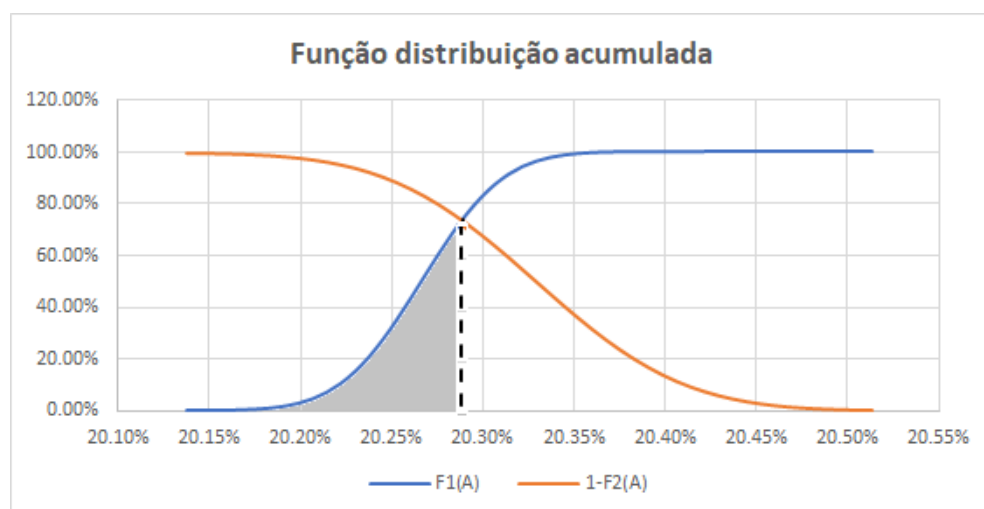


Figura 3.8: Intersecção entre $F_1(x)$ e $1-F_2(x)$

Pela figura 3.8 é possível identificar que $F_1(20.29\%) = P(X \leq 20.29\%) = 73.7\%$, ou seja, a probabilidade da "rede 1 ser melhor do que a rede 2" é de 73.7 %. No nosso caso, isto representa que a probabilidade da rede 1 ter um Mape menor do que a rede 2 é de 73.7 %.

3.6 Aplicação de redes neuronais

Após testar diferentes redes para os 3 modelos das redes neuronais, são escolhidas duas arquiteturas de redes que melhor correlacionem a temperatura ambiente e a irradiância horizontal global. Ou seja, a partir da avaliação de redes explicada na secção anterior é escolhida uma arquitectura de rede (variáveis e número de neurónios da camada escondida) da comparação de redes entre o modelo 1 e o modelo 3 e outra arquitectura de rede da comparação de redes do modelo 2. Os resultados obtidos na comparação de redes são demonstrados na secção 4.1 .

As arquiteturas com melhores resultados foram aplicadas a diferentes conjuntos de dados por forma a fazer comparações com os métodos de MCP utilizados pela Megajoule e recriar novos TMY baseados nas redes neuronais. De seguida são apresentados esses mesmos conjuntos:

- **Dados medidos de 2014 e 2015:** A aplicação das redes neuronais aos dados completos de 2014 e 2015 permitiu fazer comparações diretas com as correlações aplicadas pelos métodos de MCP utilizados pela Megajoule;
- **Dados de TMY inicial:** Aplicação das redes aos dados de TMY inicial de forma a obter um TMY correlacionado, denominado de TMY RN;
- **Dados de Longo Termo:** Aplicação das redes aos dados de Longo Termo de 2008 a 2016 por forma a construir um TMY, denominado TMY const.

3.6.1 Comparação de métodos

Após encontrar a relação entre os dados de Longo Termo e os dados medidos, as arquiteturas de redes neuronais com melhores desempenhos foram aplicadas à totalidade dos dados medidos de 2014 e 2015 de forma a fazer-se comparações com as metodologias de MCP utilizadas pela Megajoule. Nesta comparação são utilizados os indicadores de desempenho apresentados na secção 2.4.1 cujos resultados estão expostos na secção 4.2.

3.6.2 Construção do TMY

Após a aplicação da melhor rede de irradiância e de temperatura aos valores de Longo Termo, a construção do TMY const foi efetuada do seguinte modo:

- Média mensal de irradiância de cada ano (2008 a 2016);

- Comparação dos valores mensais de irradiância de cada ano com os valores mensais de irradiância da série de Longo Termo através de desvios absolutos:

$$D_i = |x_i - m(X)| \quad (3.15)$$

Nesta expressão:

- D_i é o desvio absoluto mensal de irradiância do ano i ;
 - x_i média mensal do ano i ;
 - $m(X)$ média mensal de todos os valores da série de LT.
- Escolha do valor mínimo do desvio absoluto para cada mês e inserção do ano i , associado ao desvio D_i como o ano mais "típico" desse mês;
 - O TMY fica completo com a inserção dos valores horários de temperatura simultâneos aos anos considerados no ponto anterior como "típicos" de cada mês.

3.7 Comparação do TMY

No final deste processo estão criados 4 tipos de TMY:

- **TMY inicial:** TMY construído a partir dos dados de LT iniciais, apresentados na secção 3.2.2;
- **TMY Megajoule:** TMY construído através da metodologia com melhores resultados utilizada pela Megajoule e explicada na secção 2.4;
- **TMY RN:** TMY construído através da aplicação direta da RN ao TMY inicial;
- **TMY const:** TMY construído a partir dos dados de Longo Termo correlacionados, como explicado na secção 3.6.2.

Tendo em conta os dados medidos fornecidos, foram utilizados 2 grupos de dados para comparação com os TMY criados:

- **Dados medidos de 2017:** Dados horários de irradiância horizontal global e de temperatura. Tendo em conta que estes dados não foram utilizados no processo de treino e teste da rede neuronal é importante perceber como se comportam os TMY criados;
- **Dados medidos de 2014, 2015 e 2017:** Média horária representativa de um ano dos dados de irradiância horizontal global e de temperatura da totalidade dos dados medidos. Este conjunto pretende recriar uma espécie de "TMY de dados medidos", representando provavelmente o conjunto no qual as comparações são mais importantes, tendo em conta que estamos a avaliar um conjunto maior de dados reais, e por isso, provavelmente mais representativo de um ano "típico" de valores medidos.

Por forma a avaliar e comparar os conjuntos mencionados anteriormente foram utilizados os seguintes parâmetros:

- *Mean Absolute Error* (MAE) entre cada TMY e o conjunto de dados medidos:

$$MAE = \frac{1}{N} \sum_{i=1}^N |GHI_{tmy} - GHI_{med}| \quad (3.16)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Temp_{tmy} - Temp_{med}| \quad (3.17)$$

Onde:

- $Temp_{TMY}$: temperatura horária do conjunto de TMY a avaliar;
- $Temp_{med}$: temperatura horária do conjunto de valores medidos;
- GHI_{TMY} : irradiância horária do conjunto de TMY a avaliar;
- GHI_{med} : irradiância horária do conjunto de valores medidos;

– N: número de registros.

- *Mean Absolute Percentage Error* (MAPE) entre cada TMY e o conjunto de dados medidos:

$$MAPE(\%) = \frac{1}{N} \sum_{i=1}^N \left| \frac{GHI_{tmy} - GHI_{med}}{GHI_{med}} \right| \times 100 \quad (3.18)$$

$$MAPE(\%) = \frac{1}{N} \sum_{i=1}^N \left| \frac{Temp_{tmy} - Temp_{med}}{Temp_{med}} \right| \times 100 \quad (3.19)$$

Nesta expressão:

- $Temp_{TMY}$: temperatura horária do conjunto de TMY a avaliar;
 - $Temp_{med}$: temperatura horária do conjunto de valores medidos;
 - GHI_{TMY} : irradiância horária do conjunto de TMY a avaliar;
 - GHI_{med} : irradiância horária do conjunto de valores medidos;
 - N: número de registros.
- Quantificador de ganho em relação ao modelo de referência (Imp_{MAE}^{ref}). Corresponde a um parâmetro adicional para quantificar o ganho da utilização de um determinado TMY em detrimento do modelo de referência, o TMY inicial:

$$Imp_{MAE}^{ref}(\%) = \frac{MAE^{ref} - MAE^{tmy}}{MAE^{ref}} \times 100 \quad (3.20)$$

Nesta expressão:

- MAE^{ref} : MAE do conjunto de referência (TMY inicial);
- MAE_{tmy} : MAE do conjunto de TMY a avaliar;

Capítulo 4

Exposição de resultados

Neste capítulo são apresentados e analisados os resultados provenientes da implementação das etapas explicadas no capítulo da metodologia. São também comparados entre si os diferentes modelos de TMY criados, para além da comparação com o TMY de referência, que corresponde ao TMY inicial.

4.1 Redes neuronais

De forma a clarificar a exposição de resultados apresentada para as redes neuronais, de agora em diante as variáveis utilizadas serão descritas do seguinte modo:

X1- Hora do dia	Y1- Irradiância Horizontal Global Medida
X2- Semana do ano	Y2- Temperatura Medida
X3- Irradiância de Longo Termo	Y3- Index de claridade Medido
X4- Temperatura de Longo Termo	
X5- Index de claridade de Longo Termo	

4.1.1 Modelo 1

Para o modelo 1 do treino de rede neuronal, em que o *target* é a irradiância horizontal global medida, após a adição das variáveis uma a uma as redes com melhores performances foram as seguintes:

Tabela 4.1: Melhores redes encontradas com a utilização do *Forward Method* para o modelo 1.

	Rede 1 (X1,X2,X3,X4,X5)	Rede 2 (X1,X3,X4,X5)
Mape_médio	20,06%	20,16%
Desv_padrão	0,09%	0,05%

No processo contrário, em que inicialmente todas as variáveis foram colocadas no treino da rede e retiradas uma a uma, as redes com melhores performances foram as seguintes:

Tabela 4.2: Melhores redes encontradas no *Backward method* para o modelo 1.

	Rede 3 (X1,X3,X4,X5)	Rede 4 (X1,X3,X4)
Mape_médio	20,15%	20,21%
Desv_padrão	0,04%	0,06%

De seguida, foram aumentados os neurónios da camada escondida das duas melhores redes de cada método, Rede 1 e Rede 3, por forma a encontrar a melhor arquitectura de rede possível.

Tabela 4.3: Aumento de neurónios das redes anteriores com melhores performances, N representa o número de neurónios

	Rede 1			
	N:20	N:30	N:40	N:50
Mape_médio	19,83%	19,79%	19,77%	19,79%
Desv_padrão	0,07%	0,05%	0,07%	0,06%
	Rede 3			
	N:20	N:30	N:40	
Mape_médio	20,03%	19,97%	20,00%	
Desv_padrão	0,06%	0,05%	0,05%	

Prob A>B (A melhor que B)

A \ B ->	Rede 1 (N:10)	Rede 3 (N:10)	Rede 1 (N:20)	Rede 1 (N:30)	Rede 1 (N:40)	Rede 1 (N:50)	Rede 3 (N:20)	Rede 3 (N:30)	Rede 3 (N:40)
Rede 1 (N:10)	50.0%	76.1%	7.9%	3.1%	3.1%	3.7%	42.2%	25.5%	34.2%
Rede 3 (N:10)	23.9%	50.0%	0.3%	0.0%	0.0%	0.0%	11.1%	2.2%	5.6%
Rede 1 (N:20)	92.1%	99.7%	50.0%	38.8%	32.8%	37.1%	93.8%	87.2%	91.4%
Rede 1 (N:30)	96.9%	100.0%	61.2%	50.0%	41.2%	46.9%	98.4%	95.7%	97.6%
Rede 1 (N:40)	96.9%	100.0%	67.2%	58.8%	50.0%	55.3%	98.3%	95.9%	97.5%
Rede 1 (N:50)	96.3%	100.0%	62.9%	53.1%	44.7%	50.0%	97.8%	94.8%	96.8%
Rede 3 (N:20)	57.8%	88.9%	6.2%	1.6%	1.7%	2.2%	50.0%	27.9%	39.6%
Rede 3 (N:30)	74.5%	97.8%	12.8%	4.3%	4.1%	5.2%	72.1%	50.0%	62.9%
Rede 3 (N:40)	65.8%	94.4%	8.6%	2.4%	2.5%	3.2%	60.4%	37.1%	50.0%

Figura 4.1: Comparação das redes com melhores performances do modelo 1

Da análise da figura 4.1 podemos verificar que a única rede, utilizando o processo explicado na secção 3.5.5 da metodologia, em que a probabilidade de ser melhor do que as restantes redes é sempre maior do que 50 % é a rede 1 com a utilização de 40 neurónios na camada escondida da rede neuronal.

4.1.2 Modelo 2

O modelo 2 corresponde ao modelo utilizado para encontrar a correlação entre os dados de Longo Termo e os dados medidos de temperatura.

Utilizando o processo *Forward Method*, as redes com melhores performances foram as seguintes:

Tabela 4.4: Melhores redes encontradas no *Forward method* para o modelo 2.

	Rede 5 (X2,X3,X4,X5)	Rede 6 (X1,X2,X3,X4,X5)
Mape_médio	4,02%	3,99%
Desv_padrão	0,01%	0,02%

Utilizando o processo contrário, o *Backward Method*, as melhores redes foram:

Tabela 4.5: Melhores redes encontradas no *Backward method* para o modelo 2.

	Rede 7 (X2,X3,X4,X5)	Rede 8 (X1,X2,X3,X4)
Mape_médio	4,01%	3,98%
Desv_padrão	0,02%	0,02%

Após encontrar as melhores redes dos dois processos já citados, procedeu-se ao aumento dos neurónios da camada escondida:

Tabela 4.6: Aumento de neurónios das redes anteriores com melhores performances do modelo 2, N representa o número de neurónios

	Rede 6	
	N:20	N:30
Mape_médio	3,98%	3,99%
Desv_padrão	0,02%	0,03%
	Rede 5	
	N:20	N:30
Mape_médio	3,99%	4,04%
Desv_padrão	0,01%	0,02%
	Rede 8	
	N:20	N:30
Mape_médio	3,98%	4,03%
Desv_padrão	0,02%	0,02%

É possível verificar que ao aumentar o número de neurónios da rede para 30, houve um aumento do MAPE, pelo que decidiu-se não aumentar mais este número. Posteriormente, procedeu-se à comparação de todas as redes mencionadas até agora, cujos resultados estão exibidos na seguinte figura:

Prob A>B (A melhor que B)

	Rede 6 (N:10)	Rede 5 (N:10)	Rede 8 (N:10)	Rede 5 (N:20)	Rede 5 (N:30)	Rede 6 (N:20)	Rede 6 (N:30)	Rede 8 (N:20)	Rede 8 (N:30)
Rede 6 (N:10)	50.0%	75.5%	33.5%	54.7%	88.0%	33.6%	48.3%	33.0%	85.7%
Rede 5 (N:10)	24.5%	50.0%	9.1%	24.9%	74.5%	11.7%	27.1%	11.3%	69.5%
Rede 8 (N:10)	66.5%	90.9%	50.0%	74.4%	96.2%	47.0%	61.4%	46.4%	95.4%
Rede 5 (N:20)	45.3%	75.1%	25.6%	50.0%	89.0%	26.9%	44.3%	26.3%	86.6%
Rede 5 (N:30)	12.0%	25.5%	3.8%	11.0%	50.0%	5.5%	15.2%	5.3%	43.8%
Rede 6 (N:20)	66.4%	88.3%	53.0%	73.1%	94.5%	50.0%	62.0%	49.5%	93.4%
Rede 6 (N:30)	51.7%	72.9%	38.6%	55.7%	84.8%	38.0%	50.0%	37.6%	82.4%
Rede 8 (N:20)	67.0%	88.7%	53.6%	73.7%	94.7%	50.5%	62.4%	50.0%	93.7%
Rede 8 (N:30)	14.3%	30.5%	4.6%	13.4%	56.2%	6.6%	17.6%	6.3%	50.0%

Figura 4.2: Comparação das redes com melhores performances do modelo 2

A única rede com probabilidade maior ou igual a 50 % de ser melhor do que as restantes é a rede 8 com 20 neurónios na camada escondida da rede. Assim sendo, foi esta a arquitectura de rede escolhida como melhor para correlacionar a temperatura.

4.1.3 Modelo 3

Para o modelo 3, o processo de combinação de entradas foi repetido pelo que as redes com melhores performances na primeira abordagem, referente à introdução de uma variável de entrada de cada vez estão apresentadas na seguinte tabela:

Tabela 4.7: Melhores redes encontradas no *Forward method* para o modelo 3.

	Rede 9 (X1,X3,X4,X5)	Rede 10 (X1,X2,X3,X4,X5)
Mape_médio	20,01%	19,82%
Desv_padrão	0,03%	0,06%

Quando as entradas são retiradas uma a uma as arquiteturas com melhor performance são as seguintes:

Tabela 4.8: Melhores redes encontradas no *Backward method* para o modelo 3.

	Rede 11 (X1,X3,X4,X5)	Rede 12 (X1,X2,X4,X5)
Mape_médio	20,01%	20,04%
Desv_padrão	0,03%	0,07%

De forma similar ao procedido para o modelo 1, foram aumentados o número de neurónios da camada escondida das redes com melhor performance no modelo 3 :

Tabela 4.9: Aumento de neurónios das redes anteriores com melhores performances do modelo 3, N representa o número de neurónios

	Rede 9		Rede 10		
	20 neurónios	30 neurónios	20 neurónios	30 neurónios	40 neurónios
Mape_médio	19,95%	19,95%	19,75%	19,71%	19,71%
Desv_padrão	0,05%	0,05%	0,07%	0,05%	0,06%

Comparando as diferentes redes, obtem-se a seguinte tabela de probabilidades:

Prob A>B (A melhor que B)

	Rede 9 (N:10)	Rede 10 (N:10)	Rede 9 (N:20)	Rede 9 (N:30)	Rede 10 (N:20)	Rede 10 (N:30)	Rede 10 (N:40)
Rede 9 (N:10)	50.0%	1.9%	20.9%	23.9%	0.6%	0.0%	0.0%
Rede 10 (N:10)	98.1%	50.0%	88.4%	89.7%	27.2%	13.8%	15.5%
Rede 9 (N:20)	79.1%	11.6%	50.0%	53.4%	4.4%	0.7%	1.1%
Rede 9 (N:30)	76.1%	10.3%	46.6%	50.0%	3.9%	0.6%	0.9%
Rede 10 (N:20)	99.4%	72.8%	95.6%	96.1%	50.0%	37.4%	38.3%
Rede 10 (N:30)	100.0%	86.2%	99.3%	99.4%	62.6%	50.0%	50.3%
Rede 10 (N:40)	100.0%	84.5%	98.9%	99.1%	61.7%	49.7%	50.0%

Figura 4.3: Comparação das redes com melhores performances do modelo 3

Neste caso, a rede com melhor performance é a rede 10 com 30 neurónios na camada escondida da rede.

4.1.4 Comparação modelo 1 e modelo 3

Tendo em conta que dentro destes modelos só foi escolhida uma rede, representativa da correlação da irradiância, a rede escolhida foi a rede 10 do modelo 3 por apresentar valores mais baixos de MAPE por comparação com a melhor rede do modelo 1, rede 1.

4.2 Comparação entre métodos

De forma a avaliar a qualidade da correlação obtida pelas redes neuronais, a melhor arquitectura de redes de temperatura ambiente e de irradiância foi testada para os valores totais de LT de 2014 e 2015, obtendo-se dados correlacionados de irradiância e de temperatura dos valores de 2014 e 2015.

Posteriormente, estes dados são comparados com os dados correlacionados a partir dos métodos de MCP utilizados na Megajoule. Os parâmetros utilizados na comparação foram descritos na secção 3.6.1, pelo que a tabela seguinte representa os resultados da aplicação desses parâmetros aos dados correlacionados:

Tabela 4.10: Resultados dos parâmetros utilizados para a comparação entre métodos de MCP

	GHI GAMAS	GHI MOD	GHI RN	Temp RN	Temp Megajoule
MBE	-1.05 %	0.38 %	0.13 %	0.11 %	3.23 %
R²	0.9124	0.9114	0.9265	0.9377	0.8760
MAPE	22.11 %	21.84 %	19.98 %	3.58 %	6.24 %

Conclui-se, portanto, que a correlação efetuada pelas redes neuronais entre os valores de Longo Termo e os valores medidos de 2014 e 2015 apresenta melhores resultados, tendo em conta os parâmetros de análise utilizados para a avaliação. Os valores de MBE, R² e MAPE têm valores mais baixos para a irradiância e para a temperatura. Importante referir que tendo em conta o MBE mais baixo, o método de irradiância utilizado pela Megajoule escolhido para as comparações de TMY, foi o método GHI MOD, uma vez que os restantes parâmetros de erro são bastante próximos em relação ao método GHI GAMAS.

4.2.1 Construção do TMY

Como explicado na secção 3.6.2 decidiu-se aplicar as arquiteturas das redes neuronais escolhidas aos valores de Longo Termo, obtendo-se assim um *output* de valores de Longo Termo correlacionados.

O processo para obtenção do TMY const a partir dos valores de Longo Termo correlacionados, foi explicado na secção 3.6.2. A figura 4.4 representa a média mensal de irradiância de todos os anos de Longo Termo utilizados.

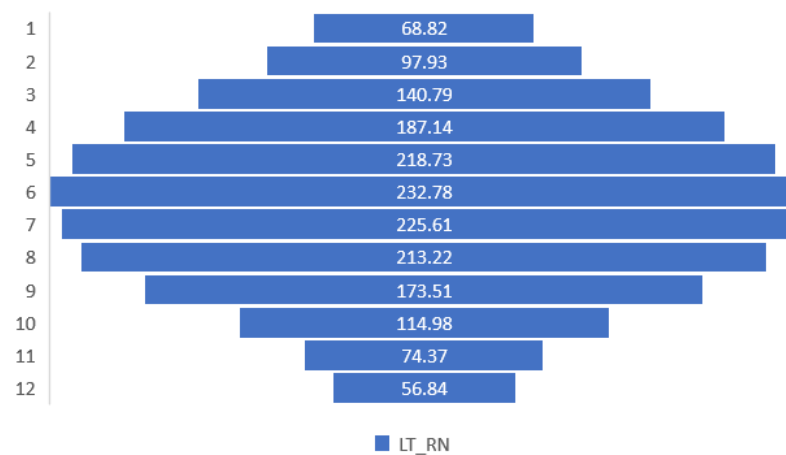


Figura 4.4: Média mensal de irradiância dos dados de Longo termo correlacionados (W/m^2)

De seguida, a média mensal de cada mês de cada ano foi comparada com a média mensal total calculada anteriormente. A figura 4.5 representa os desvios absolutos mínimos encontrados para cada mês do ano:

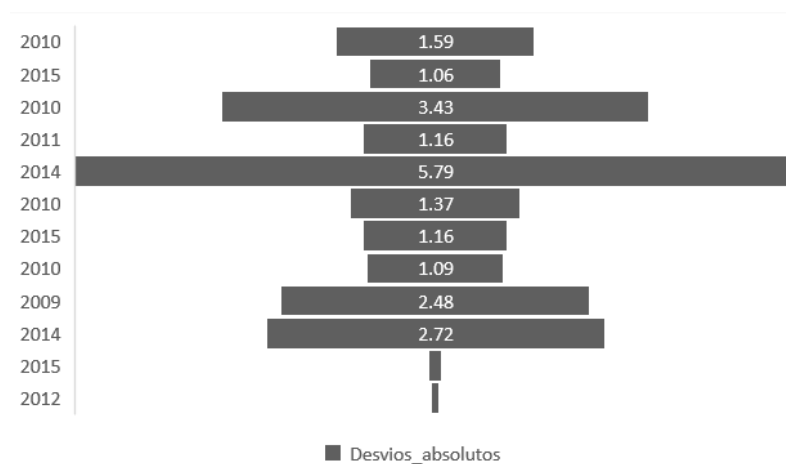


Figura 4.5: Desvios absolutos em relação à média para cada mês (W/m^2). Eixo dos yy's representa o ano cujo o desvio absoluto desse mês foi menor

Posto isto, a compilação de anos utilizada no TMY const foi a seguinte:

Tabela 4.11: Ano inserido no TMY const para cada mês e desvio absoluto em relação à média mensal de Longo Termo

Mês	Ano utilizado	Desvio absoluto(W/m^2)
1	2010	1.59
2	2015	1.06
3	2010	3.43
4	2011	1.16
5	2014	5.79
6	2010	1.37
7	2015	1.16
8	2010	1.09
9	2009	2.48
10	2014	2.72
11	2015	0.10
12	2012	0.06

4.3 Comparação de TMY's com valores medidos de 2017

4.3.1 Irradiância

O erro médio mensal representado no gráfico abaixo mostra que o TMY RN está mais próximo dos valores mensais de irradiância do ano de 2017 em todos os meses do ano em relação ao TMY inicial e ao TMY Megajoule.

O TMY const tem desvios menores em relação aos restantes TMY na maior parte dos meses, ainda que se verifique um comportamento oscilatório. No mês 5, 9 e 12 verificaram-se desvios maiores do TMY const em relação ao TMY Megajoule.

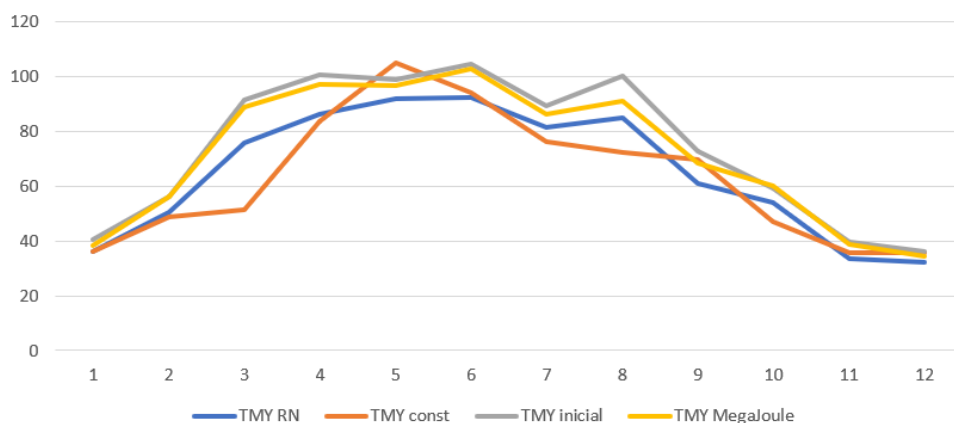


Figura 4.6: Erro médio absoluto (MAE) mensal de irradiância em relação aos valores do ano de 2017

Para todos os meses do ano, o TMY RN tem um ganho positivo em relação ao TMY inicial. O TMY const tem um ganho maior em relação aos restantes TMY na maior parte dos meses (7 meses), ainda que no mês de maio represente uma perda (-6,05 %) em relação ao TMY inicial.

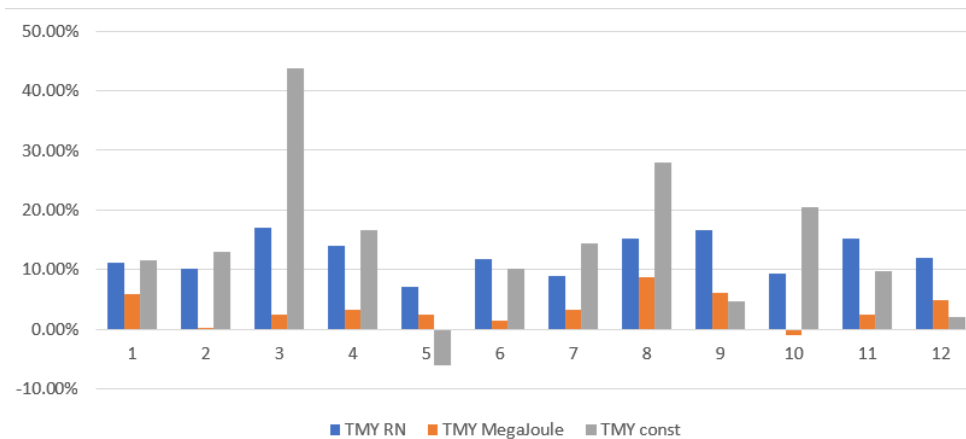


Figura 4.7: Ganho (%) em relação ao MAE mensal do modelo de referência (TMY inicial)

Do ponto de vista anual, a comparação dos TMY com o TMY inicial permite concluir que o TMY com maior ganho foi o TMY const, seguido de muito perto pelo TMY RN. Os indicadores de MAE e MAPE são, também, mais favoráveis ao TMY const, ainda que o TMY RN tenha valores muito próximos.

Tabela 4.12: Parâmetros de comparação de TMY's

	TMY RN	TMY const	TMY inicial	TMY Megajoule
IMP_inicial (%)	12.42%	15.13%	—	3.47%
MAE (W/m ²)	65.00	62.99	74.22	71.65
MAPE(%)	43.93%	42.57%	50.16%	48.42%

4.3.2 Temperatura

O TMY const tem valores de MAE mensais menores na maior parte dos meses (6 meses), ainda que no mês de janeiro e abril seja o modelo com maior desvio em relação aos valores de temperatura de 2017.

O mês com desvios mais altos é o mês de outubro, no qual em média os modelos de TMY têm um desvio de 2,94°C, o que poderá querer dizer que o mês de outubro de 2017 foi um mês com temperaturas desfasadas em relação ao normal.

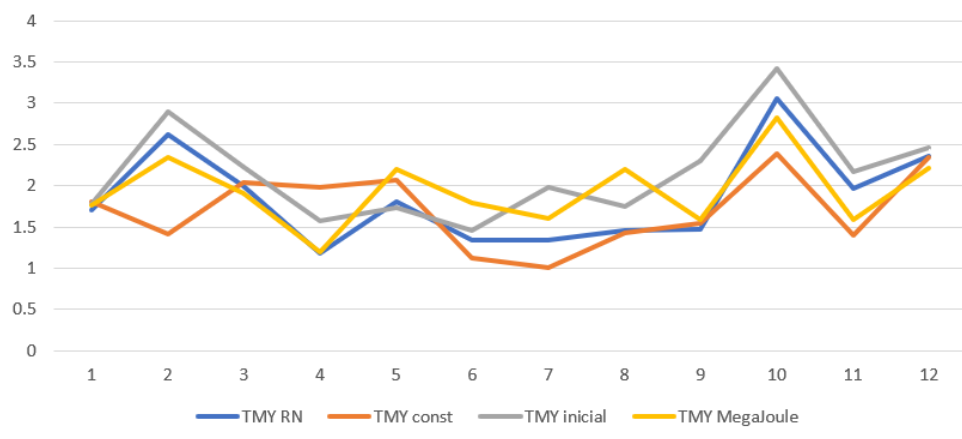


Figura 4.8: Erro médio absoluto (MAE) mensal de temperatura em relação aos valores do ano de 2017

O TMY RN teve um ganho superior em relação ao TMY inicial na maior parte dos meses, à exceção do mês de maio no qual houve uma perda de (3,92%). O TMY const teve uma perda em relação ao TMY inicial nos meses de janeiro, abril e maio, sendo o modelo com ganho maior em 6 meses do ano.

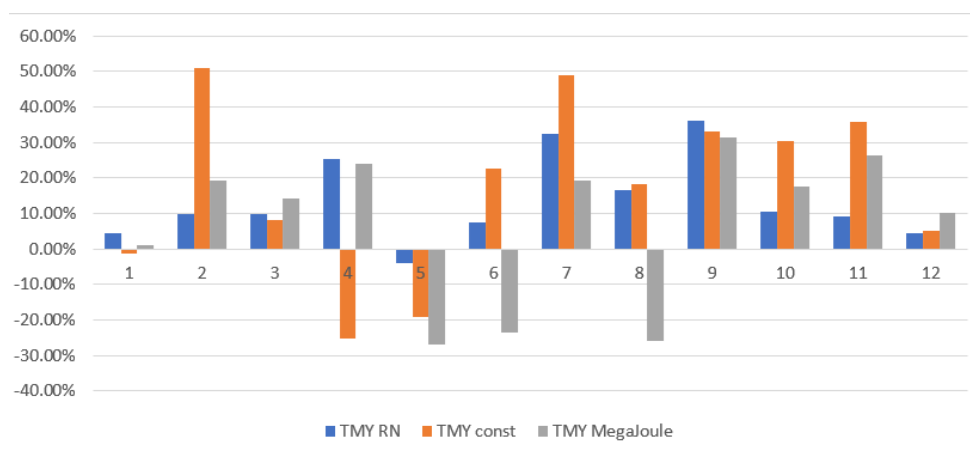


Figura 4.9: Ganho (%) em relação ao MAE mensal do modelo de referência (TMY inicial)

O modelo de TMY com melhores parâmetros anuais foi o TMY construído, no qual houve um ganho global de 19,97 % em relação ao TMY inicial e cerca de 10 % em relação ao TMY utilizado pela Megajoule.

Tabela 4.13: Parâmetros de comparação de TMY's

	TMY RN	TMY const	TMY inicial	TMY Megajoule
IMP_inicial (%)	13.40%	19.97%	—	9.70%
MAE (°C)	1.86	1.72	2.14	1.94
MAPE(%)	10.05%	9.28%	11.60%	10.48%

4.4 Comparação de TMY's com média da totalidade de valores medidos

4.4.1 Irradiância

O TMY const é o modelo com MAE mais baixo para todos os meses, à exceção do mês de setembro e de dezembro. O TMY RN tem uma aproximação melhor em todos os meses por comparação com o TMY Megajoule e o TMY inicial.

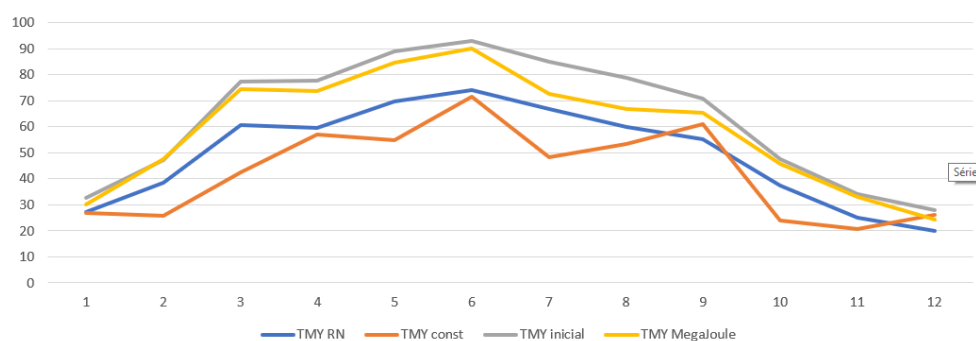


Figura 4.10: Erro médio absoluto (MAE) mensal de irradiância em relação à média da totalidade dos valores medidos

Há exceção de setembro e de dezembro, o ganho mensal do TMY const em relação ao TMY inicial é superior a todos os outros modelos. Por outro lado, o TMY Megajoule representa o modelo com ganhos menores em relação ao TMY inicial em todos os meses, à exceção do mês de dezembro.

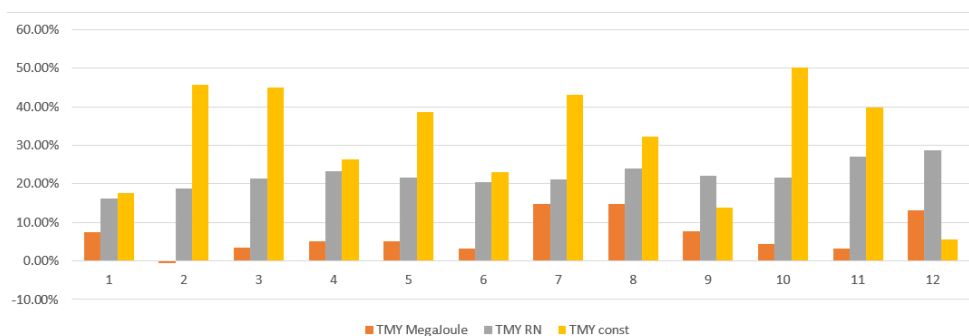


Figura 4.11: Ganho (%) em relação ao MAE mensal do modelo de referência (TMY inicial)

Os parâmetros globais de comparação dos TMY são favoráveis ao modelo TMY const. De salientar o ganho de 32,80 % desse modelo em relação ao TMY inicial e de mais de 25 % em relação ao ganho do método de MCP da Megajoule. Os parâmetros do TMY RN demonstram um ganho de 21,98 % em relação ao TMY inicial e cerca 15 % em relação ao TMY utilizado pela Megajoule.

Tabela 4.14: Parâmetros de comparação de TMY's

	TMY RN	TMY const	TMY inicial	TMY Megajoule
IMP_inicial (%)	21.98%	32.80%	—	6.98%
MAE (W/m^2)	49.55	42.68	63.51	59.07
MAPE (%)	33%	28.43%	42.30%	37.35%

4.4.2 Temperatura

O TMY const tem MAE menores em relação aos restantes modelos em 8 dos 12 meses do ano. Há exceção do mês de março, os modelos de TMY const e de TMY RN foram os modelos com menores MAE mensais.

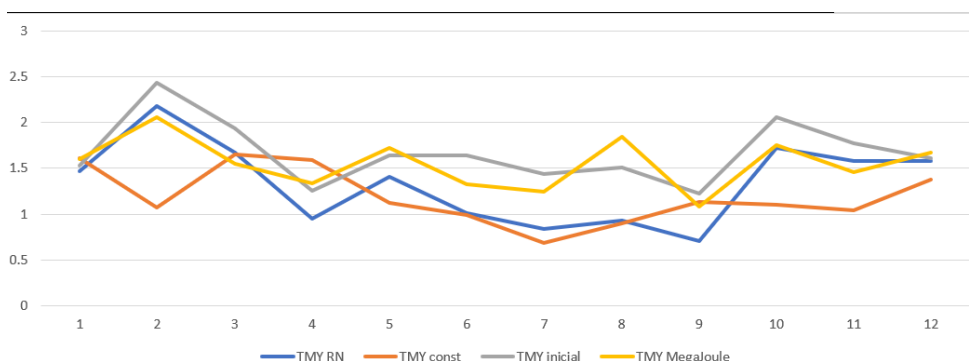


Figura 4.12: Erro médio absoluto (MAE) mensal de temperatura em relação à média da totalidade dos valores medidos

O modelo TMY const representa o modelo com maior ganho em relação ao TMY inicial em 8 dos 12 meses. No mês de janeiro e de abril, o TMY const tem uma perda em relação ao TMY inicial de (-5,42%) e (-26.60%), respectivamente. O TMY RN representa o único modelo com ganhos para todos os meses em relação ao TMY inicial.

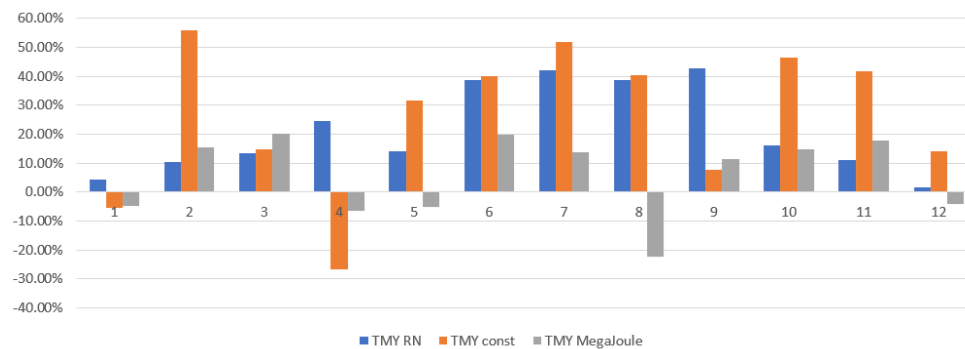


Figura 4.13: Ganho (%) em relação ao MAE mensal do modelo de referência (TMY inicial)

Para o conjunto total de dados, o modelo com melhores parâmetros é o modelo TMY const representando um ganho de 28,59 % em relação ao TMY inicial, mais de 20% em relação ao ganho proporcionado pelo TMY Megajoule. Os dois modelos criados a partir de rede neurais têm portanto um melhor desempenho em comparação com o TMY dos dados medidos, representativo da média dos dados de 2014, 2015 e 2017.

Tabela 4.15: Parâmetros de comparação de TMY's

	TMY RN	TMY const	TMY inicial	TMY Megajoule
IMP_inicial (%)	20.07%	28.59%	—	6.81%
MAE (°C)	1.33	1.19	1.67	1.55
MAPE(%)	7.35%	6.57%	9.20%	8.57%

Capítulo 5

Conclusões e Trabalho Futuro

5.1 Conclusões

Os objetivos principais desta dissertação poderiam ser divididos em três partes distintas: desenvolver uma ferramenta de MCP, mais tarde decidiu-se utilizar redes neuronais, portanto neste caso o objetivo seria encontrar duas arquiteturas de rede, uma para irradiância e outra para temperatura, que pudessem correlacionar os valores de Longo Termo com os valores medidos, comparar a nova ferramenta/metodologia de MCP com as metodologias já conhecidas e por fim, construir um TMY a partir do novo método de correlação.

Para a correlação da irradiância, a rede com melhor desempenho foi a rede com as 5 variáveis de entrada do vetor de entradas escolhido e com 30 neurónios na camada escondida do modelo 3. Para a correlação da temperatura, a rede com melhor desempenho foi a rede com 4 variáveis de entrada (irradiância de longo termo, temperatura, hora do dia e semana do ano) e com 20 neurónios na camada escondida do modelo 2.

Na comparação com os métodos conhecidos, os outputs das redes neuronais demonstraram ter uma previsão melhor, através dos parâmetros de comparação utilizados: MAPE, R e MBE.

Foram construídos dois tipos de TMY baseados nas redes neuronais: TMY const e TMY RN. Para o conjunto total de valores medidos (2014, 2015 e 2017), o TMY const demonstrou ser o TMY com melhor previsão de irradiância e de temperatura através dos parâmetros de comparação utilizados: MAE, MAPE e Imp_{MAE} . Pela análise dos mesmos parâmetros, o TMY RN demonstrou ter uma melhor aproximação à totalidade dos dados medidos por comparação com o TMY inicial e o TMY Megajoule. De referir que os valores altos de MAPE da tabela 4.12, referentes aos desvios de irradiância dos TMY's em relação aos registos do ano de 2017, poderão indicar que o ano de 2017 foi um ano "atípico" ou que o conjunto de anos utilizado no treino de rede para obter as correlações (2014 e 2015) não é representativo da tipicidade de irradiância do local.

Para este caso de estudo, conclui-se que a metodologia utilizada, baseada nas redes neuronais, é uma alternativa válida de MCP e pode ser facilmente implementada uma vez que não implica gastos adicionais.

5.2 Trabalho Futuro

A análise de comparação dos TMY construídos carece de informação de campanha de valores medidos, pelo que em trabalhos futuros seria interessante obter uma gama de valores medidos mais ampla, por forma a que as correlações efetuadas fossem mais precisas.

O facto de não haver medição de irradiância difusa neste caso de estudo , permite deixar em aberto a criação de um novo modelo de redes neuronais por forma a correlacionar a irradiância difusa.

As comparações efetuadas foram feitas apenas em função da irradiância e da temperatura, pelo que seria interessante em trabalhos futuros efetuar comparações com os valores de produção de saída do *software* PVsyst. Neste caso, por forma a melhorar o TMY const, os anos de Longo Termo correlacionados cujos anos coincidem com os dados medidos, poderiam ser substituídos pelos dados desses mesmos anos medidos, uma vez que estes dados podem ser considerados como os dados "reais", e daí construir-se um novo TMY.

Seria de grande importância a implementação desta metodologia noutros casos de estudo, de forma a perceber se haveriam mudanças significativas nas arquiteturas das melhores redes neuronais e se esta metodologia poderia ser de facto considerada uma alternativa válida de MCP no desenvolvimento dos projectos.

Referências

- [1] Jan Kleissl. *Solar Energy Forecasting and Resource Assessment*. 2013.
- [2] Sobrina Sobri, Sam Koohi-Kamali, e Nasrudin Abd. Rahim. Solar photovoltaic generation forecasting methods: A review. *Energy Conversion and Management*, 156:459–497, jan 2018.
- [3] Can Wan, Jian Zhao, Yonghua Song, Zhao Xu, Jin Lin, e Zechun Hu. Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems*, 1(4):38–46, dec 2015.
- [4] *TRENDS 2018 IN PHOTOVOLTAIC APPLICATIONS Survey Report of Selected IEA Countries between*. 1992.
- [5] Sam Koohi-Kamal, N.A. Rahim, H. Mokhlis, e V.V. Tyagi. Photovoltaic electricity generator dynamic modeling methods for smart grid applications: A review. *Renewable and Sustainable Energy Reviews*, 57:131–172, may 2016.
- [6] Cyril Voyant, Gilles Notton, Soteris Kalogirou, Marie-Laure Nivet, Christophe Paoli, Fabrice Motte, e Alexis Fouilloy. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105:569–582, may 2017.
- [7] Projecto "GREENPRO". *Energia Fotovoltaica- manual sobre tecnologias, projecto e instalação*. 2004.
- [8] Oscar Perpiñán Lamigueiro. *Energía Solar Fotovoltaica*. Relatório técnico, 2011.
- [9] Luis. Jutglar Banyeras. *Energia solar*. CEAC, 2004.
- [10] Global Solar Atlas.
- [11] A. (Antonio) Luque e Steven. Hegedus. *Handbook of photovoltaic science and engineering*. Wiley, 2003.
- [12] Abdoul Kader Sidibe, Mounir Ghogho, Chaker El Amrani, e Mustapha Faquir. A comparative study of simple clear sky irradiance models. Em *2016 International Conference on Electrical and Information Technologies (ICEIT)*, páginas 113–117. IEEE, may 2016.
- [13] Elke Lorenz, Jan Remund, Stefan C Müller, Wolfgang Traunmüller, Gerald Steinmaurer, David Pozo, José Antonio Ruiz-Arias, Vicente Lara Fanego, Lourdes Ramirez, Martin Gaston Romeo, Christian Kurz, Luis Martin Pomares, e Carlos Geijo Guerrero. BENCHMARKING OF DIFFERENT APPROACHES TO FORECAST SOLAR IRRADIANCE. Relatório técnico.

- [14] Peder Bacher, Henrik Madsen, e Henrik Aalborg Nielsen. Online short-term solar power forecasting. *Solar Energy*, 83(10):1772–1783, oct 2009.
- [15] Bella Espinar, José-Luis Aznarte, Robin Girard, Alfred Mbairadjim Moussa, Georges Kariniotakis, B Espinar, J L Aznarte, R Girard, A M Moussa, e G Kariniotakis. *Photovoltaic Forecasting: A state of the art*. 2010.
- [16] Tom Stoffel. *Terms and Definitions*. 2007.
- [17] Marius Paulescu, Eugenia Paulescu, Paul Gravila, e Viorel Badescu. *Weather Modeling and Forecasting of PV Systems Operation*, volume 103. 2013.
- [18] KS Reddy, , M Ranjan Energy Conversion Management, e undefined 2003. Solar resource estimation using artificial neural networks and comparison with other correlation models. *Elsevier*.
- [19] Amit Kumar Yadav e S.S. Chandel. Solar radiation prediction using Artificial Neural Network techniques: A review. *Renewable and Sustainable Energy Reviews*, 33:772–781, may 2014. doi:10.1016/J.RSER.2013.08.055.
- [20] Honglu Zhu, Xu Li, Qiao Sun, Ling Nie, Jianxi Yao, Gang Zhao, Honglu Zhu, Xu Li, Qiao Sun, Ling Nie, Jianxi Yao, e Gang Zhao. A Power Prediction Method for Photovoltaic Power Plant Based on Wavelet Decomposition and Artificial Neural Networks. *Energies*, 9(1):11, dec 2015.
- [21] Vladimiro Miranda. *Redes Neurais – treino por retropropagação*. 2007.
- [22] Atika Qazi, H. Fayaz, A. Wadi, Ram Gopal Raj, N.A. Rahim, e Waleed Ahmed Khan. The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. *Journal of Cleaner Production*, 104:1–12, oct 2015.
- [23] R. Aggarwal e Y. Song. Artificial neural networks in power systems. Part 1: General introduction to neural computing. *Power Engineering Journal*, 11(3):129–134, jun 1997.
- [24] Maria Helena, Osório Pestana, e Engenharia Electrotécnica. Departamento de Engenharia Electrotécnica e de Computadores Avaliação e Controlo de Segurança de Redes Interligadas com Grande Penetração Eólica com base em Métodos de Aprendizagem Automática. 2007.
- [25] Christian A. Gueymard. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renewable and Sustainable Energy Reviews*, 39:1024–1034, nov 2014.
- [26] Ali Dinler. A new low-correlation MCP (measure-correlate-predict) method for wind energy forecasting. *Energy*, 63:152–160, dec 2013.
- [27] José A. Carta, Sergio Velázquez, e Pedro Cabrera. A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site. *Renewable and Sustainable Energy Reviews*, 27:362–400, nov 2013.
- [28] Krystian Cieslak e Piotr Dragan. Comparison of the existing photovoltaic power plant performance simulation in terms of different sources of meteorological data.
- [29] MathWorks - Makers of MATLAB and Simulink - MATLAB & Simulink. URL: <https://www.mathworks.com/?s{ }tid=gn{ }logo>.

- [30] Tai-cong Chen, Da-jian Han, F.T.K. Au, e L.G. Tham. Acceleration of Levenberg-Marquardt training of neural networks with variable decay rate. Em *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 3, páginas 1873–1878. IEEE.
- [31] Muriel Gevrey, Ioannis Dimopoulos, e Sovan Lek. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160(3):249–264, feb 2003.