

Definição de limiares de aceitação para porcentagem de nuvem por Redes Neurais Artificiais para controle de qualidade de dados geoespaciais - estudo de caso de imagens LandSat

Tatiana Pará Monteiro de Freitas¹
Alex de Lima Teodoro da Penha¹
Carlos Frederico de Sá Volotão¹

¹Instituto Militar de Engenharia -IME. Departamento de Engenharia Cartográfica.
Praça Gen. Tibúrcio, 80 - Urca - 22290-270. Rio de Janeiro - RJ. Brasil.
tatianapara@hotmail.com, alexdapenha@gmail.com, volotao@ime.eb.br

Abstract. In Brazil, is still difficult to find a standard for the use of remote sensing images, as there is a climatic heterogeneity inherent of the Brazilian territory that generate clouds and prevent the viewing of large areas of Earth's surface. Therefore, it was adopted the use of artificial neural networks (ANN) as a technique to try to solve this problem. This article proposes a methodology based on artificial neural networks for predicting percentage of clouds in a given period of the year. It used the metadata of 354 images (between Landsat 5 and 7), both the input data and the output of the simulated neural network. Also has been used nine input data (date of passage in Julian days, revolution, north latitude, west longitude, south latitude, east longitude, Greenwich Mean Time (GMT), the solar azimuth, and elevation of the sun) and the outputs are related to the percentage cloud of each quadrant of the image obtained by its respective metadata: Q1; Q2; Q3; and Q4. The architecture of the ANN was tested at various levels and the topology 9-28-28-8-4 was the one that got the best statistical result with an error in the expected results of + - 2.78%, consistent with the complexity of the problem. Therefore, it is possible to solve the problem of prediction of percentage of cloud in a specific region with an error of + - 3%.

Palavras-chave: Remote Sensing, neural Networks, Clouds, Sensoriamento Remoto, Redes Neurais, Nuvens.

1. Introdução

Os avanços tecnológicos recentes e o aumento de usuários de informações geográficas (IG) tem proporcionado um crescimento exponencial do comércio de dados geoespaciais nos últimos anos, principalmente dos dados matriciais. Mediante este fato a corrida pela geração e aquisição de dados de Sensoriamento Remoto (SR) fica cada vez mais veloz, fazendo com que os produtores de IG busquem tomar medidas de qualidade para a aquisição e disponibilização dos produtos finais (p.ex: Imagem satélite que opera na faixa do visível). Mas um aspecto do controle de qualidade de dados geoespaciais refere-se à omissão de dados (Lunardi, 2012). Por isso conhecer previamente a possibilidade de ocorrência de nuvens fruto das condições meteorológicas inerentes a área de estudo permite um planejamento de projetos para aquisição de dados de SR dentro de várias limitações de coleta de dados (como cronograma, orçamento e escopo).

A ISO (*International Organization for Standardization*) e conseqüentemente a Infraestrutura Nacional de Dados Espaciais - INDE (BRASIL, 2008; CONCAR, 2010) prevê que a porcentagem de nuvem seja definida de forma determinística no controle de qualidade de produtos e insumos cartográficos. A medida de aceitação da presença de nuvem em um determinado dado geoespacial de uma determinada região passa a ser relevante, principalmente quando se trata de uma grande densidade de dados em extensas áreas. Tapakis (2013) apresenta uma detalhada revisão bibliográfica sobre o assunto.

Mas a dificuldade está em prever e identificar graus de confiabilidade em que se aceita o risco de omissão de dados nestas fontes, até o qual sejam considerados desnecessários e aceitáveis para um determinado projeto e região. Adiciona-se a isso que a alta resolução temporal dos produtos de satélite é inversamente proporcional a escala e ao valor do custo de aquisição, de tal forma que aumentar a possibilidade de se obter imagens sem nuvem pode significar o uso de um produto de pior resolução espacial, para imagens de aquisição gratuita,

como ocorre, por exemplo, com os dados do sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS) (NASA, 2011) com resolução temporal diária e resolução espacial de 250 x 250m, limitado para o mapeamento moderno com cada vez mais resolução espacial (Rudorff et al, 2007).

Para estudar aspectos não conhecidos da realidade, têm-se cada vez mais utilizado técnicas de Redes Neurais Artificiais (RNA) que usam de um método de aprendizagem inteligente para demonstrar a realidade, ou seja, simulam o comportamento dos neurônios biológicos no cérebro. A figura a seguir (figura 1) apresenta a topologia de uma rede neural.

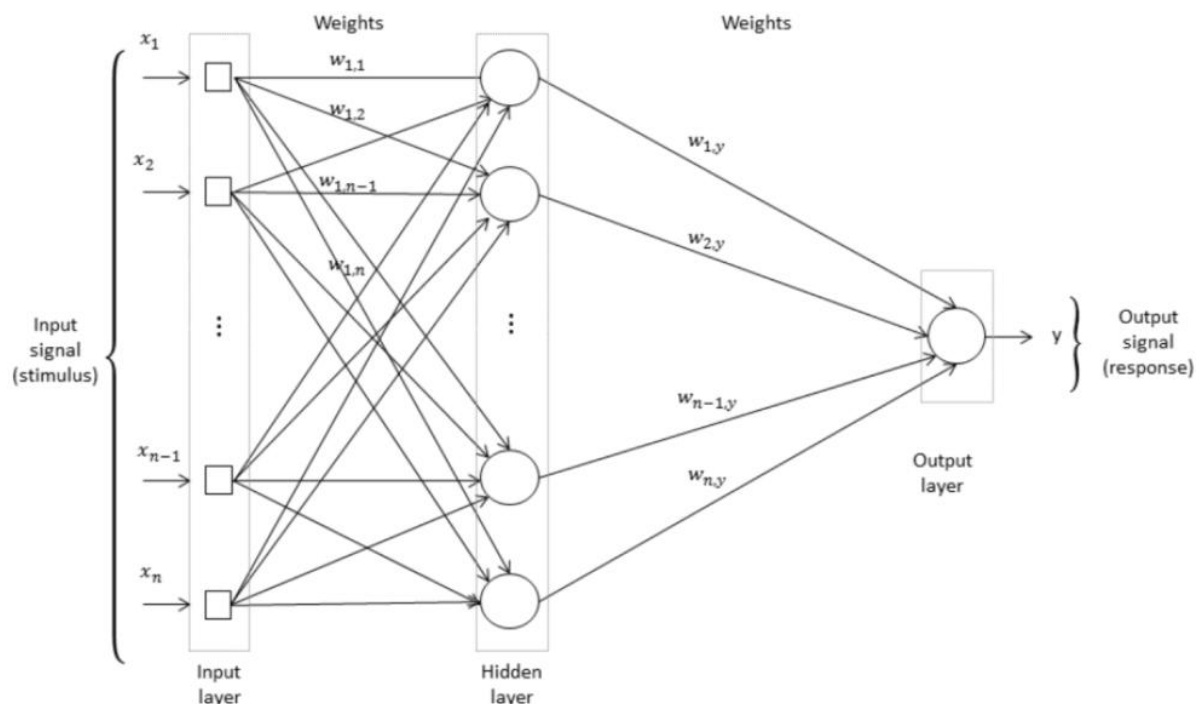


Figura 1. Topologia de rede neural *feedforward* (sem realimentação). x_n representa a n -ésima unidade de entrada, e y representa a unidade de saída. (Saponaro, 2013)

Atualmente ainda há poucas metodologias de controle de qualidade para aplicação em projetos de grande porte, sendo em geral definidas e testadas em pequenas áreas, como ensaio piloto no âmbito acadêmico e não para produção em larga escala (Santos et al., 2008).

Uma parte importante da pesquisa de qualidade de dados espaciais refere-se a descrição do erro e incerteza em dados espaciais (Oort, 2005). Por isso, pretende-se testar uma metodologia baseada em redes neurais artificiais para previsão de porcentagem de nuvens em um determinado período do ano que possa ser aplicada a grandes áreas de mapeamento, construindo um estimador neural capaz de prever os limites de aceitação no valor de conformidade máxima de porcentagem de nuvem em cada período do ano para cada região.

A metodologia e os resultados apresentados a seguir representam os primeiros processamentos referentes a este tipo de pesquisa.

2. Metodologia de Trabalho

Coletou-se dados do Sensor "Mapeamento Temático" (*Landsat Thematic Mapper - TM*) presentes nos produtos do LandSat 5, e dados de imagens do LandSat 7 sensor ETM. Os parâmetros coletados foram os presentes nos metadados das imagens se satélite LandSat, como também a percentagens de cobertura de nuvens dos 4 quadrantes presentes na imagem satélite (Q1, Q2, Q3 e Q4), todos para o período entre os anos de 1993 até 2011, da região que compreendida pela Cena LandSat com órbita e ponto 216/069. Todos editados em uma tabela

e processadas no software *open source* Mbp (*Multiple Back-Propagation*) para formação de redes neurais.

Nos primeiros testes foram usados 14 dados (oriundos dos metadados) de entradas e 4 saídas (Q's) em um total de 324 amostras e 30 para teste. Na etapa seguinte, após a identificação das entradas mais relevantes e de, conseqüentemente, um entendimento melhor da complexidade do problema, utilizou-se apenas as entradas consideradas relevantes e determinísticas para o treinamento da rede neural como sendo apenas 9: data de passagem em dias Julianos, revolução, latitude norte, longitude oeste, latitude sul, longitude leste, tempo central (GMT), azimute do sol, e elevação do sol. Para os dados de teste de treinamento da rede foram sorteados aleatoriamente no conjunto total de dados. A tabela a seguir (tabela 1) apresenta os dados que fizeram parte das simulações finais.

Tabela 1. Dados usados para testes finais: 9 dados de entrada e 4 dados de Saída.

Data	Revol.	Lat. N	Long. O	Lat. S	Long. L	Tempo	Az. Sol	E. Sol	Q1	Q2	Q3	Q4
8	47106	-1.20961	-4.03194	-1.3928	-3.89177	12.07	109.47	512.372	90	80	80	70
24	47339	-1.20969	-4.03103	-1.39272	-3.89028	12.07	104.559	502.768	100	80	100	90
40	47573	-1.20963	-4.03277	-1.39279	-3.89206	12.07	975.262	497.737	90	80	90	90
56	47806	-1.20963	-4.02978	-1.39281	-3.88955	12.07	889.204	494.775	90	40	80	70
104	48504	-1.20965	-4.02969	-1.39281	-3.89031	12.08	611.669	46.256	100	90	100	80
120	48738	-1.20961	-4.03254	-1.39282	-3.89253	12.08	54.166	439.952	100	100	90	90
152	49204	-1.20963	-4.03394	-1.39281	-3.89306	12.08	464.957	394.326	40	90	10	90
168	49437	-1.20963	-4.03019	-1.39279	-3.88967	12.08	457.457	380.414	90	80	90	70
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
106	44270	-1.21054	-4.02663	-1.39193	-3.88671	12.53	546.695	517.156	90	100	100	100
122	44503	-1.21054	-4.02645	-1.39191	-3.88625	12.53	475.563	489.177	90	90	90	90
154	44969	-1.21051	-4.02795	-1.39193	-3.88703	12.53	405.424	438.019	100	100	80	90
186	45435	-1.21048	-4.02762	-1.39193	-3.88711	12.53	415.043	423.062	90	70	80	60
202	45668	-1.2105	-4.02715	-1.39194	-3.88706	12.53	441.573	435.536	10	30	90	20
218	45902	-1.21049	-4.02668	-1.39194	-3.88654	12.52	479.961	460.991	100	100	100	100
266	46601	-1.21058	-4.0261	-1.39184	-3.88525	12.52	679.395	580.534	80	80	80	70
282	46834	-1.21069	-4.02401	-1.39174	-3.88318	12.52	785.633	614.238	90	80	90	80
314	47300	-1.21053	-4.02414	-1.39193	-3.88382	12.5	101.979	633.195	100	100	100	100

Ressalta-se que a metodologia baseou-se nas informações dos metadados das imagens Landsat 5 e 7 para os dados de entrada e saída das redes neurais artificiais simuladas. Outros dados poderiam ser usados como temperatura da superfície, umidade do ar, velocidade e direção do vento, evaporação, etc. Mas este tipo de dado de entrada implicaria em previsões apenas de curto prazo, visto que o objetivo não está na caracterização da ocorrência e detecção de nuvem e sim à possibilidade de ocorrência em uma determinada área. Adiciona-se o fato das propriedades das nuvens serem muito variáveis.

A figura 2 apresenta o fluxograma da metodologia aplicada no processamento de imagens Landsat e implementada por redes neurais artificiais.

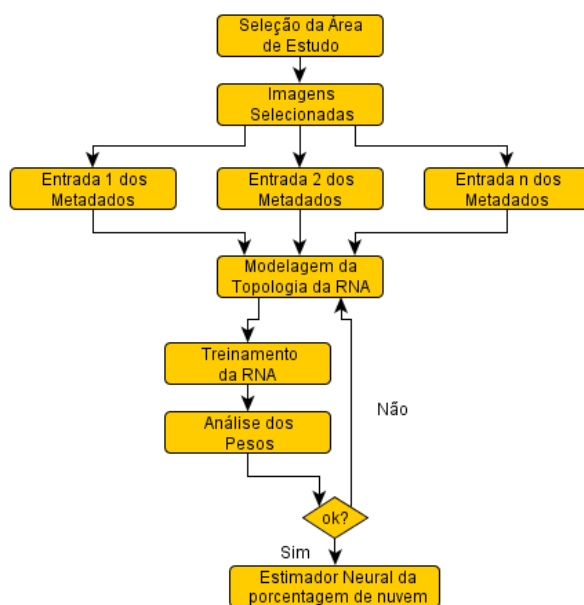


Figura 2. Metodologia aplicada no processamento de imagens Landsat e implementada por redes neurais artificiais.

A capacidade de aprendizado de uma rede neural parece ser adequada a este tipo de problema. Ainda, como não há uma regra que defina o número de camadas ou o número de neurônios na rede, foram ainda testadas as topologias: 14-10-4, 14-28-4, 14-7-2-4, 14-7-7-4, 14-28-8-4, 14-10-10-4, 14-8-16-6-4, 9-8-16-6-4 e 9-28-28-8-4.

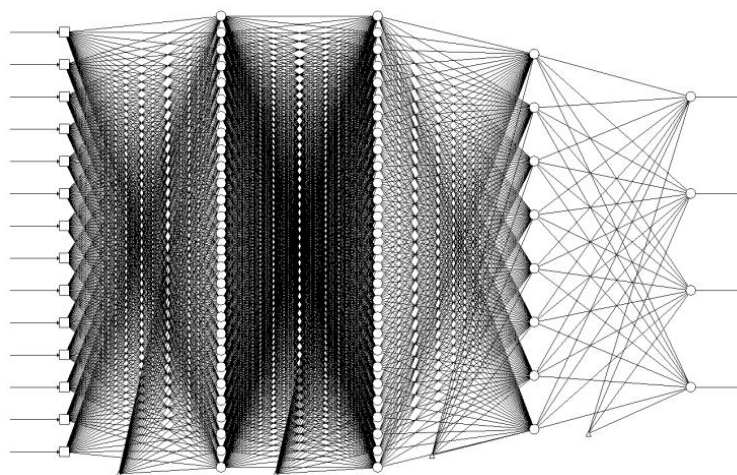


Figura 3. Rede neural com uma arquitetura de 14 entradas e 4 saídas, sendo 3 camadas de neurônios, de 28, 28 e 8, respectivamente.

Alguns autores consideram diferença aceitável aquela que for menor ou igual a um erro de 1% ou 0,5%, em média (Belugo, 2002). Para este trabalho considerou-se aceitável estimar 3% como aceitável, dada a complexidade do problema, a partir do EMQ e com um padrão de parada em 3.000.000 de épocas.

3. Resultados e Discussão

Nos gráficos a seguir (figura 4) são apresentados os resultados estatísticos dos dados de treinamento e teste, para Q1 a partir da topologia adotada para a rede.

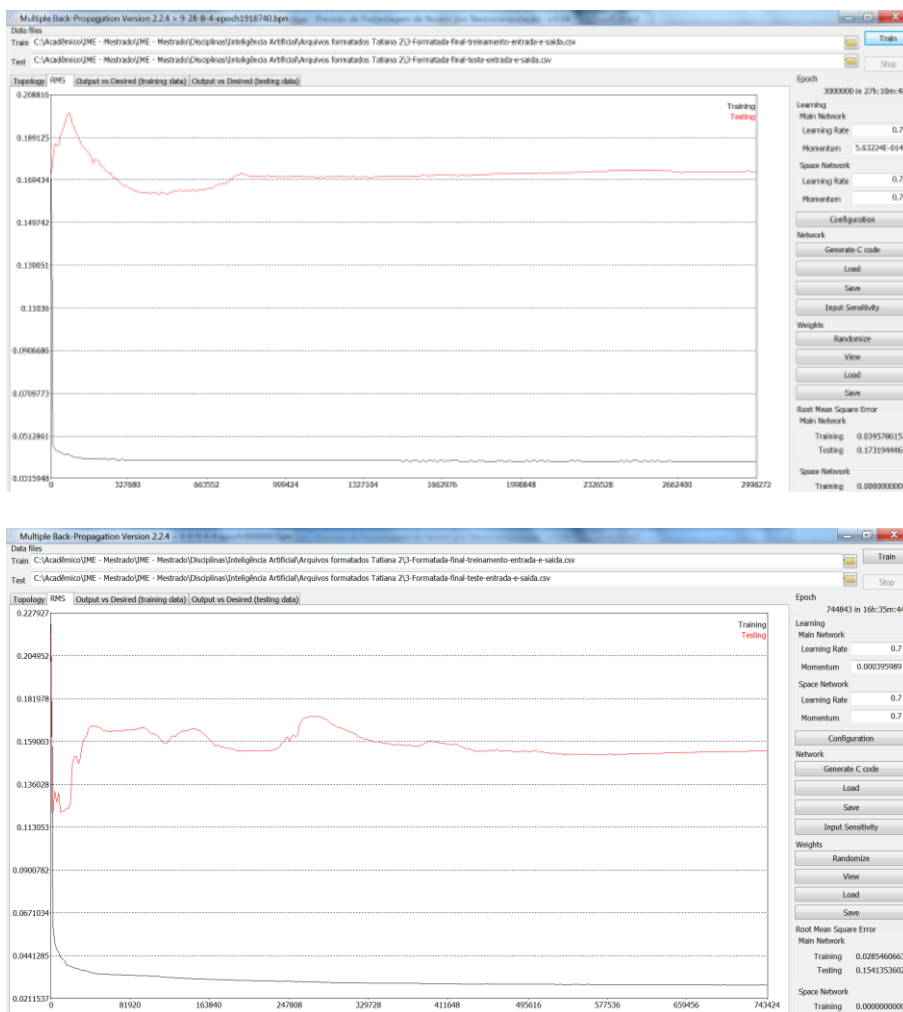


Figura 4. Topologia 9-28-8-4 e Topologia 9-28-28-8-4, respectivamente.

Os resultados estatísticos são apresentados na tabela a seguir (tabela 2).

Tabela 2. Resultados estatísticos dos testes entre as duas topologia de redes com melhores resultados.

Topologia da Rede	Erro do Treinamento	Erro do Teste	Tempo
9-8-16-6-4	0.0500989377	0.0857615319	29h28m43s
9-28-28-8-4	0.0278620185	0.1653818013	32h25m20s

A topologia 9-28-28-8-4 foi a que obteve o melhor resultado estatístico, como pode ser observado nas imagens de gráficos a seguir.



Figura 5. Saída versus desejado (dados de treinamento para Q1)

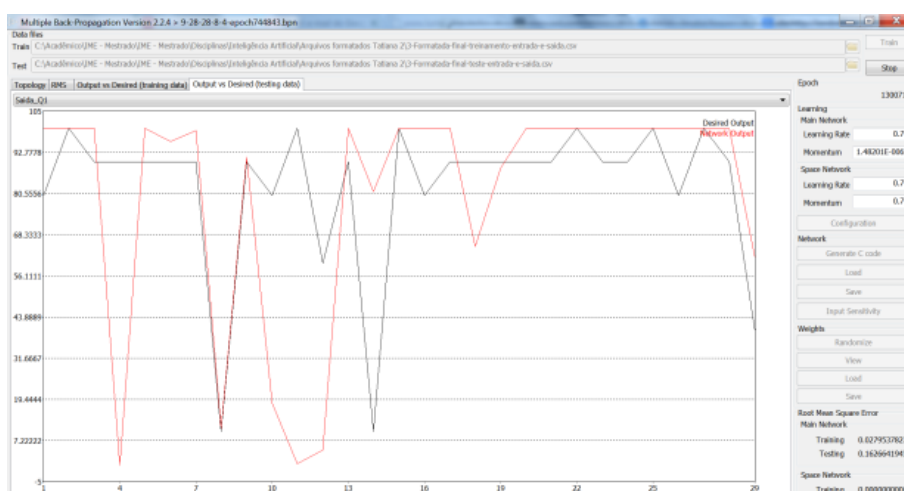


Figura 6. Saída versus desejado (dados de teste para Q1)

O estimador neural capaz de prever os limites de aceitação no valor de conformidade máxima de porcentagem de nuvem em cada período do ano é dado por matrizes de peso.

Obteve-se um erro no resultado esperado de $\pm 2.78\%$, consistente com a complexidade do problema, que habitualmente é tratado apenas com previsões de curto prazo.

4. Conclusões

O modelo neural encontrado permite uma avaliação de consistência a priori e que impacta diretamente na quantidade de áreas vazias em um conjunto de dados geoespacial tanto vetorial quanto matricial.

Trata-se da completude do dado no que diz respeito à presença de nuvens, no qual os valores encontrados permitem antecipar o risco da omissão de dados devido a presença de nuvens e, talvez, do uso de tecnologias alternativas para o mapeamento.

Conclui-se que, a partir dos resultados mostrados, é possível resolver o problema de previsão de porcentagem de nuvem em uma determinada região com um erro de $\pm 3\%$.

A porcentagem de nuvem encontrada, na prática, é um valor de conformidade máxima, ou seja, menor porcentagem de nuvem que pode ser exigida para a região em estudo em um determinado período.

Referências Bibliográficas

- Beluco, Adriano. **Classificação de Imagens De Sensoriamento Remoto Baseada Em Textura Por Redes Neurais**. 2002. 154p. Dissertação (Mestrado em Sensoriamento Remoto) - Universidade Federal Do Rio Grande Do Sul, Centro Estadual De Pesquisa Em Sensoriamento Remoto E Meteorologia - CEPSRM, Curso De Pós-Graduação Em Sensoriamento Remoto. 2002.
- BRASIL. Decreto nº 6.666, de 27 de novembro de 2008. **Institui, no âmbito do Poder Executivo federal, a Infraestrutura Nacional de Dados Espaciais - INDE, e dá outras providências**. Diário Oficial da República Federativa do Brasil, Brasília, DF, 28 nov. 2008. Seção 1, pt. 1, p. 57.
- CONCAR – Comissão Nacional de Cartografia. **Plano de ação para implantação da infraestrutura de dados espaciais (INDE)**. Rio de Janeiro , 2010.
- Lunardi, Omar Antônio; Penha, Alex de Lima Teodoro da; Cerqueira, Rodrigo Wanderley de. **O Exército Brasileiro e os Padrões de Dados Geoespaciais para a INDE**. IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação (SIMGEO), Recife, PE. 2012. ISBN: 978-85-63978-01-1
- NASA MODIS. Disponível em <<http://modis.gsfc.nasa.gov/>>. Acesso em 14 de julho de 2013
- Oort, P. V. **Spatial data quality: from description to application**. Delf: NCG, Nederlandse Commissie voor Geodesie. Publications on geodesy, 0165-1706; 60. ISBN 9061322952. 2005.
- Rudorff, B.F.T.; Shimabukuro, Y.E.; Ceballos, J.C. **Sensor MODIS e suas Aplicações Ambientais no Brasil**. 1. ed. São José dos campos: Parêntese, 2007. 425p.
- Santos, A.G. DOS ; Segantine, P. C. L. ; Silva, I. DA . **Controle de qualidade de mapas digitais urbanos para uso em Sistemas de Informações Geográficas**. Revista Minerva, v. 5, p. 169-176, 2008.
- Saponaro, G., Kolmonen, P., Karhunen, J., Tamminen, J., And de Leeuw, G.: **A neural network algorithm for cloud fraction estimation using NASA-Aura OMI VIS radiance measurements**. Atmospheric Measurement Techniques (Journal Atmos. Meas. Tech),. 6, 2301-2309. Doi: 10.5194/amt-6-2301-2013. 2013
- Tapakis, R., Charalambides, A.G. **Equipment and methodologies for cloud detection and classification: A review**. Solar Energy. V.95, 392-430, Doi: 10.1016/J.Sol.Energy-392-2013. 2013