Cloud top Height Estimation from the Meteosat Water Vapor Imagery Using Computational Intelligence Techniques: Preliminary Results

Stavros Kolios, Petros Karvelis, Chrysostomos Stylios
Department of Informatics and Telecommunication Technology
Technological Educational Institute of Epirus
Arta, Greece
e-mail: stavroskolios@yahoo.gr, stylios@teleinfom.teiep.gr, pkarvelis@gmail.com

Periklis Tagkas
Department of Applied Foreign Languages in Management and Commerce
Technological Educational Institute of Epirus
Igoumenitsa, Greece
e-mail: ptangas@teiep.gr

***Abstract*—This study investigates the cloud top height estimation using nonlinear methods to Meteosat imagery. The suggested approach aims to develop an integrated statistical methodology to estimate the cloud top height on a pixel basis using Meteosat Second Generation water vapor imagery. Radiosonde measurements are used as reference dataset and a spatio-temporal correlation with Meteosat images is performed in order to collect a representative sample for the statistical analysis. Here, we apply Multi Layer Perceptron (MLP) and Support Vector Machines (SVM) and we compare the results to the Linear Regression model. The best results are achieved using SVM for regression. The proposed approach is very promising as it can be used for future in-depth analysis so as to develop a robust approach for geometrical height estimation on a pixel basis of the operational data of Meteosat imagery. It is noted that an accurate estimation of cloud top height can help to eliminate geometric restrictions (e.g. Parallax phenomenon) of the Meteosat satellite imagery, improving its usefulness in a wide area of applications and especially in satellite-based weather forecast.**

Keywords-Regression; SVM; MLP; Meteosat; Radiosondes; Geometric height

I. INTRODUCTION

Geostationary satellites nowadays provide many different data about the parameters concerning land, ocean and atmosphere and they are used in numerous applications not only in detecting but also in monitoring (on a real-time basis) and short-range forecasting [1], [2].

There are some limitations because of the geostationary satellite instrument field of view and their scanning mode that usually lead to inaccuracies about the exact location of some cloud features (especially convective cloud systems) on the Earth's surface. These cloud systems are usually vertically extended at several kilometers in the atmosphere, with the cloud tops reaching the middle and upper troposphere [3-5]. The above-mentioned inaccuracies are caused by the parallax phenomenon, [6] which is more evident in the convective nature cloud systems that are depicted at satellite images and appears at mid-latitudes growing towards the poles and the edges of the field of view of the satellite instruments.

The convective cloud systems have great importance because they are related to extreme weather phenomena and events like lightning, hail, strong winds and floods [5], [7] and [8].

As a result, the accurate estimation of Cloud Top Height (CTH) is appropriate in order to calculate and eliminate the parallax effect and it can provide the real location of the convective systems above the Earth's surface, more accurately.

There are several studies that propose different approaches and methodologies to estimate the geometric CTH but they are mainly lacking in high accuracy or usually refer to small geographic areas [9-12].

Our effort is to propose a robust, generalized and accurate methodology to estimate the geometrical height of different cloud features (especially convective ones) at a broad range of latitudes and longitudes. This methodology aims to improve on pixel basis the accuracy of clouds' location above the Earth's surface, so as to eliminate the parallax phenomenon in real-time monitoring applications based on Meteosat operational data.

Section II provides all the relative information about the data used and the area of interest. In Section III, the methodology was followed in order to estimate cloud top height, is provided. The Section IV presents the accuracy results of the proposed methodology and the Section V provides the conclusions of the the study.

II. DATA AND AREA OF INTEREST

Two different datasets were used in this study. Firstly, radiosonde observations (provided by the National Oceanic and Atmospheric Administration/Earth System Research Laboratory: NOAA/ESRL) of the greater area of the Mediterranean region (Fig. 1). Radiosonde data are suitable to measure vertical profiles of many important parameters like temperature, pressure and geometrical height. Secondly, Meteosat satellite images of two different channels (5 and 6) at spectral ranges of 5.35 - 7.15 μm (spectral center: 6.2 μm) and 6.85 - 7.85 μm (spectral center: 7.3 μm) respectively

 ISBN: 978-1-61208-264-6

(provided via the European Organization for the Exploitation of Meteorological Satellites: EUMETSAT). It is mentioned that the specific Meteosat channels have been chosen because they belong to the water vapor absorption spectral region and are used to detect water vapor levels especially in the middle and upper troposphere. All the data used were collected for a period between 5/6/2012 and 14/6/2012. The area of interest covers the greater area of the Mediterranean basin (Fig. 1) because in this region the convective systems are developing frequently and especially during the warm period of the year [13] and [14] causing floods, strong winds and hail among others that affect importantly humans and properties. These cloud systems are vertical extended in several kilometers.

The accurate detection of their spatial extent on the Earth's surface of the field of view of a geostationary satellite like Meteosat, needs an accurate estimation of the CTH in order to eliminate the parallax phenomenon that is evident in such convective cloud systems [6], [12].

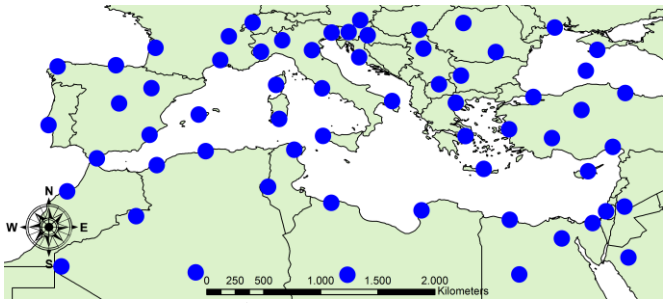

Figure. 1. Locations of radiosonde stations that are used in this study.

III. METHODOLOGY

A representative training sample was needed in order to determine the non-linear functions that combine the Meteosat pixel values with the cloud geometrical height that is recorded by the radiosonde measurements (Fig. 2). For this reason, a spatiotemporal correlation between radiosonde measurements and satellite images was implemented.

More specifically, the radiosonde balloons are usually released twice a day at times 00:00 UTC and 12:00 UTC. For the selected time period, a database was gathered with all the satellite images that have 90-min time distance from the point where radiosonde balloons were released. The 90-min time gap was chosen because our intention is to estimate optically thick and vertically extended cloud systems that are related to severe weather. These are generally called convective cloud systems where the cloud tops existed usually in the middle or the upper troposphere. The balloons reach these atmospheric levels at about 90 minutes after their release according to [15]. So, for each satellite image obtained 90 minutes after the radiosonde balloon release, a 5 x 5 kernel window (~25 x 25 km according to the Meteosat image pixel size at mid-latitudes) was created and implemented around the coordinates of every available Radiosonde point (Fig. 3). The mean Brightness Temperatures (BT) value of the two used channels and their mean value of BT differences of the kernel window pixels, are finally associated with the radiosonde measurement that was considered just above the cloud top (Fig. 2). In order to determine, if the radiosonde passed through a cloud and which is the height of its cloud top, we used the criterion of Relative Humidity level as it was proposed in [16]. If the value of the RH was higher than 85% (maximum value is the saturation level: 100%) then we consider that the balloon was in a cloud. More analytically, the higher (in km) measurement of a radiosonde balloon recordings that has RH lower than 85% and at least the three previous and consecutive measurements with RH higher than 85%, was considered as cloud top measurement (Fig. 2).

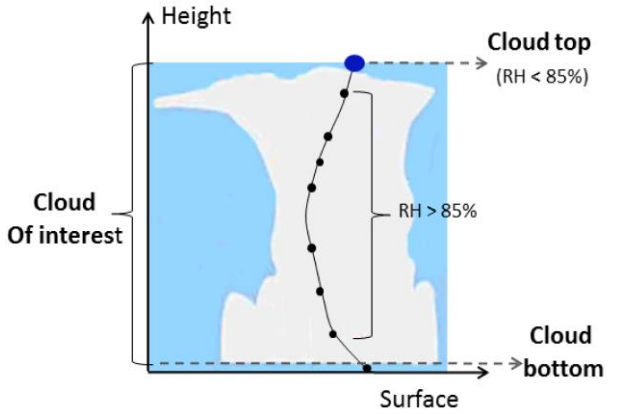

Figure 2. Schematic representation of a representative cloud system and its cloud top height, as it is calculated with the proposed methodology.

It is mentioned that the size of the kernel window was chosen because the radiosonde balloons during their ascent can reach more than 20 km horizontal distance from their initial location as it was mentioned in [15].

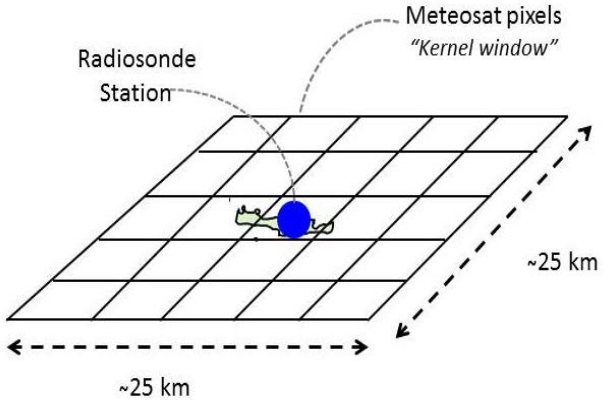

Figure 3. The kernel window of Metosat pixels that was implemented around the location of a radiosonde station.

The spatio-temporal correlation considering all the above-mentioned criteria was implemented in an automated

way using an algorithm developed for this purpose in the VB.NET programming language.

Using the above mentioned spatio-temporal correlation methodology for the radiosonde and Meteosat data, a final dataset of 181 cases in the area of interest for the selected time period, was collected. Each case is a given radiosonde dataset that satisfies the RH criteria. For each case, the geometric height above the ground was recorded (as it was measured from the radiosonde) along with the mean BT in the channels 5 and 6 of the kernel window Meteosat pixels and their mean BT difference. These samples (of 181 cases) comprised our final data set for the regression analysis.

More specifically, we defined our dataset as $\{x_i^1, x_i^2, x_i^3, y_i\}, i = 1,\ldots,n$ where n is the number of measured samples (181 cases), the x_i^1 is the mean BT value of Meteosat kernel window pixels in the channel 5, x_i^2 is the mean BT value of Meteosat kernel window pixels in the channel 6, x_i^3 is the BT difference between them and y_i is the measured height of the balloon for each one of the collected cases.

This is a classical regression problem where the variable that we have to predict is y_i. Here, we examine a Linear Regression model and two non-linear methods: Multi-Layer Perceptron (MLP) Neural Network and Support Vector Machines (SVM) , in order to predict this value.

A. Linear Regression

A linear regression model assumes that the relationship between the dependent variable y_i and the p-vector of observed x_i is linear. Thus, the model has the general form:

$$\mathbf{Y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}\,, \tag{1}$$

where

$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_p \end{pmatrix}^T$: are the regression coefficients,

$\mathbf{X} = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^p \\ x_2^1 & x_2^1 & \cdots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^p \end{pmatrix}$: is the observation matrix and

$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_p \end{pmatrix}^T$:are the error terms.

The parameters $\boldsymbol{\beta}$ are estimated using the Least Squares method and a standard solution is obtained [4].

B. MLP

Neural networks are used as a direct substitute for multivariable regression and other statistical analysis techniques. A neural network is used to represent a nonlinear mapping between input and output vectors [2, 18].

Multi-Layer Perceptron (MLP) neural network is the most widely used neural network architecture for classification and/or regression problems. MLP networks consist of an input layer, one or more hidden layers and an output layer (Fig. 4 represents an MLP architecture with one hidden layer). Each layer has a number of units and each unit is fully interconnected with weighted connections to units in the subsequent layer.

The output of MLP is defined as a linear combination of the outputs of the hidden layer nodes where every neuron uses a weighted average of the inputs through a function (e.g. sigmoid function).

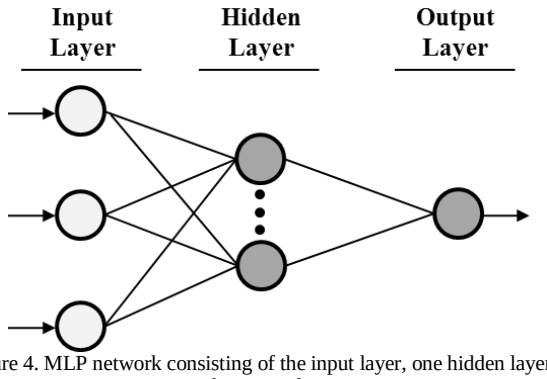

Figure 4. MLP network consisting of the input layer, one hidden layer and the output layer.

The MLP transforms p inputs to l outputs through some nonlinear functions. The output of the network is determined by the activation of the units in the output layer as follows:

$$g_0 = f(\sum_h g_h w_{h0})\,, \tag{2}$$

where $f()$ is the activation function, g_h is the activation of h -th hidden layer node and w_{h0} is the weight of the connection between the h -th hidden layer node and 0-th output layer node. The most used activation function is the sigmoid and it is given as follows:

$$g_0 = \frac{1}{1+e^{-\sum_h g_h w_{h0}}}\,. \tag{3}$$

The activation level of the nodes in the hidden layer is determined in a similar fashion. The error function is defined as the difference between the calculated value and the target value, as follows:

$$E = \frac{1}{2}\sum_i^n \sum_0^l (y_0^{(i)} - g_0^{(i)})\,, \tag{4}$$

where n is the number of pattern in data set and l is the number of output nodes.

The aim is to reduce the error by adjusting the interconnections between layers. The weights are adjusted using the Back Propagation (BP) algorithm. The algorithm

 ISBN: 978-1-61208-264-6

requires a training data that consists of a set of corresponding input and target pattern values y_i . During the training process, MLP starts with a random set of initial weights and then training continues until the set of w_{ih} and w_{h0} are optimized so that a predefined error threshold is met between g_0 and y_0 [3].

C. SVM for regression

Support Vector Machines are applied not only to classification problems but also to regression ones. SVM contain all the main features that characterize maximum margin algorithms: a nonlinear function is learned by linear learning machine mapping into high dimensional kernel induced feature space.

In SVM regression, the input x is firstly mapped onto a m-dimensional feature space using nonlinear mapping, and then a linear model is constructed in this feature space. The linear model is given by:

$$f(x,w)=\sum_{j=1}^{m} w_j g_j(x)+b\,, \qquad (5)$$

where $g_j(x),\ j=1,\ldots,m$ denotes a set of nonlinear transformations and b is the bias term.

SVM regression uses a new type of loss function called ε-insensitive loss function based on [5,6]:

$$L_\varepsilon(y, f(x,w))=\begin{cases} 0, & \text{if } |y-f(x,w)|\le\varepsilon \\ |y-f(x,w)|-\varepsilon, & \text{otherwise} \end{cases}. \qquad (6)$$

SVM regression performs linear regression in the high-dimension feature space using ε-insensitive loss and tries to reduce model complexity by minimizing $\|w\|^2$. This can be described by introducing slack variables $\xi_i^*, \xi_i, i=1,\ldots,n$, to measure the deviation of training samples outside ε - zone.

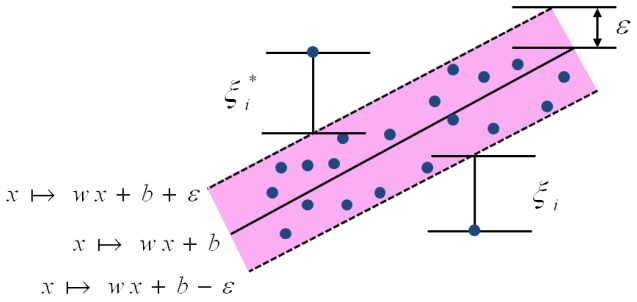

Figure 5. Epsilon zone with slack variables and selected data points.

Thus SVM regression is formulated as minimization of the following functional [6]:

$$\text{minimize} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i+\xi_i^*)$$

$$s.t. \quad \begin{cases} y_i - f(x_i,w) \le \varepsilon+\xi_i^*\,, \\ f(x_i,w)-y_i \le \varepsilon+\xi_i^* \\ \xi_i^*, \xi_\iota \ge 0, \iota=1,\ldots,n \end{cases} \qquad (7)$$

IV. RESULTS

We have tested and evaluated the above-mentioned regression models using a dataset of $n=181$ samples. 80% of the dataset was randomly selected and used as the train set and the rest (20%) of the dataset was used as test set.

TABLE I. THE AVAILABLE DATASET DIVIDED INTO TRAIN AND TEST SET.

	Number of Instances (%)	**Number of Instances (#)**
Train Set	80	145
Test Set	20	36
Data Set	**100**	**181**

A. MLP architecture

In order to determine the best architecture for the MLP Neural Network, we tested different formulations consisting of one or two hidden layers, with 3 up to 20 neurons in each one. Table II, presents the best architectures for one and two hidden layers respectively in terms of R-coefficient and Mean Absolute Error.

TABLE II. BEST MLP ARCHITECTURES FOR ONE AND TWO HIDDEN LAYERS.

Layers	**Architecture (Input-Hidden-output) nodes**	**Correlation coefficient**	**Mean Absolute Error (MAE)**
One Hidden Layer	3-18-1	0.7272	1541.05
Two Hidden Layers	3-18-2-1	0.7372	1399.21

B. SVM parameter selection

In order to suggest the most appropriate values for the parameters C , ε of SVM, we applied a GridSearch method [7]. Table III, presents the results in terms of the correlation coefficient and MAE for the determined parameters that reported with the maximum correlation coefficient.

TABLE III. SVM PERFORMANCE FOR THE PARAMETERS $C=25$, $\varepsilon=0.1$.

Layers	**Correlation coefficient**	**Mean Absolute Error (MAE)**
SVM	0.7238	1286.97

C. Comparison of the Regression models

Table IV presents a comparison of the results of the three models used in our experiments including the Linear Regression method.

TABLE IV. COMPARISON OF THE THREE MODELS.

Layers	Correlation coefficient	Mean Absolute Error (MAE)
Linear Regression	0.7200	1401.93
MLP	**0.7372**	1399.21
SVM	0.7238	**1286.97**

Fig. 6 illustrates the SVM predicted values and the actual values for the test set.

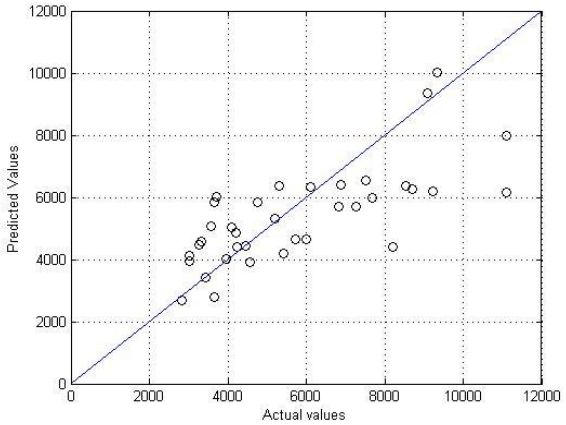

Figure 6. The SVM predictions and the actual values.

Fig. 7 displays the error of the predicted height versus the actual height in meters for the test set. As one can observe as the altitude rises from 8000 to 11000 km, the error becomes significantly large.

This is mainly due to the lack of a sufficient number of cloud tops (as they are defined with the RH criterion) in heights between 8000 and 1200 km in the selected time period. This fact led to a small number of collected cases and consequently to significant errors of the regression procedures at these heights. Increasing the time period of study, the number of the convective cases in the study area and consequently the number of convective cloud tops above 8000m will be increased. This goes beyond the scope of this study and it will be our first priority in the our future work.

In this study, all three recorded features were used in order to estimate the height; it would be interesting to study the effect of each one. For this reason we conducted the same experiments with all the combinations of the features and we present the results in terms of the correlation coefficient and MAE in Table V.

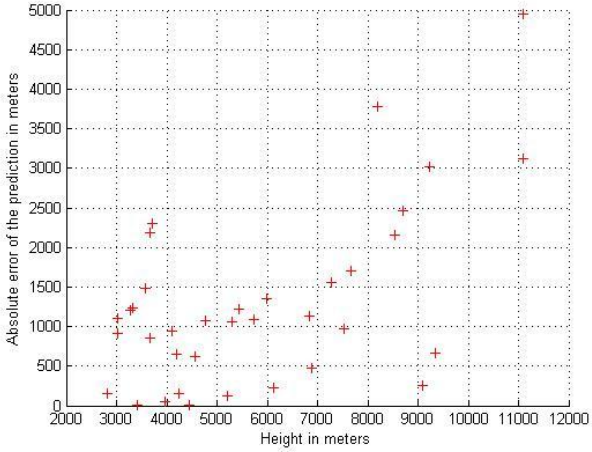

Figure 7. Absolute error of the prediction versus height (in meters).

It is concluded that the most powerful feature is the mean BT value of Meteosat kernel window pixels in channel 5, then is the BT difference between channels 5 and 6 and last is the mean BT value of Meteosat kernel window pixels in channel 6.

TABLE V. SVM REGRESSION USING EACH FEATURE.

	$BT_{channel\ 5}$	$BT_{channel\ 6}$	$BT_{channel\ 5} - BT_{channel\ 6}$
Corr. Coef.	**0.6979**	0.3348	0.4017
MAE	**1524.3**	1975.2	1813.6

V. CONCLUSIONS AND FUTURE WORK

In this study, some preliminary results regarding the cloud height estimation from Meteosat water vapor imagery were presented. An analytical and automated spatio-temporal methodology is suggested to investigate the correlation of radiosonde measurements and Meteosat Brightness Temperature values. For the estimation of geometrical cloud top height, three methodologies were examined: a linear one and two nonlinear algorithms, MLP and SVM.

Our purpose is to gradually develop a robust and accurate algorithm for CTH estimation for different cloud features (and especially the convective ones) as they are depicted in Meteosat infrared imagery. Our final intention is to cover the operational needs of CTH estimation in large geographic areas, with a simple, accurate and easy to use dataset. Using accurate estimations of the CTH, we intend in the near future to estimate (and reduce) the parallax phenomenon which is evident in the geostationary orbit satellite imagery, mainly used in the satellite-based weather forecasting.

The first results showed that mean errors vary in general between 1000 m and 1500 m and there are significant errors especially after the height of 8000 m in the atmosphere. These errors could be explained due to the lack of data for height above 8000 m in the selected time period. These findings require further investigation which is beyond the scope of this study. Nevertheless, the correlation coefficients

 ISBN: 978-1-61208-264-6

were considered quite satisfactory and promising for all the used algorithms and their parameterizations.

Future work will concern a categorization of the different types of clouds in order to study separately every discrete category of the cloud features that are depicted in the Meteosat satellite imagery. The handling of high cloud tops that penetrate the lower stratosphere is an additional topic that we are planning to study, too.

What is of great importance also, is the collection of a larger dataset that will allow more robust results and the use of additional channels of Meteosat imagery in the infrared region in order to evaluate their performance in the geometric height estimations.

In addition to this, other approaches such as Random Forests [18], could be used to try to test and improve the accuracy of the cloud height estimations.

ACKOWLEDGMENT

This work is partially supported by the project "Adriatic Port Community System APC" (code 111) funded by IPA Adriatic Cross-border Cooperation Programme 2007-2013 co-financed by European Union, Instrument for Pre-Accession Assistance.

REFERENCES

[1] A. D. Vila, A. L. Machado, H. Laurent, and I. Velasco, "Forecast and Tracking the Evolution of cloud clusters (FORTRACC) using satellite infrared imagery: methodology and verification," Weather Forecast. 2008, vol. 23, pp. 233-245.

[2] S. Kolios and H. Feidas, "An automated nowcasting system of Mesoscale Convective Systems for the Mediterranean basin using Meteosat imagery. Part I: System description," Meteorological Applications, 2012, DOI: 10.1002/met.1282.

[3] S. P. Ray, "Mesoscale Meteorology and Forecasting," American Meteorology Society, Boston, 1986.

[4] A. Roger and Sr. Pielke, "Mesoscale Meteorological Modeling," International geophysics series, Academic Press, Second Edition, 2002.

[5] A. R. Houze S. A. Rutledge, M .I. Biggerstaff, and B. F. Smull, "Interpretation of Doppler weather-radar displays in mid-latitude mesoscale convective systems," Bulletin of American Meteorological Society, 1989, vol. 70, pp. 608-719.

[6] G. A. Vicente, J. C. Davenport, and R. A. Scofield, "The role of orographic and parallax corrections on real time high resolution satellite rainfall rate distribution" International Journal of Remote Sensing, 2002, vol. 23, pp. 221-230.

[7] J. F. Correoso, E. Hernandez, R. Garcia-Herrera, E. Barriopedro, and D. Paredes, "A 3-year study of cloud-to-ground lightning flash characteristics of MCS over Western Mediterranean Sea," Atmospheric Research, 2006, vol., pp. 79 89-107.

[8] A. L. Machado, W. Lima, O. Pinto and A. Morales Relationship between cloud-to-ground discharge and penetrative clouds: A multi-channel satellite application," 2008, Atmospheric Research, In press.

[9] G. G. Campell and K. Holmlund, "Geometric cloud heights from Meteosat," International Journal of Remote Sensing, 2004, vol. 25, pp. 4505-4519.

[10] C. M. Naud, B.A. Baum, M. Pavolonis, A. Heidinger, R. Frey, and H. Zhang, "Comparison of MISR and MODIS cloud-top heights in the presence of cloud overlap," Remote Sensing of Environment, 2007, vol. 107, pp. 200-210.

[11] R. Borde and P. Dubuisson, "Cloud top height estimation using simulated METEOSAT-8 radiances", Proc. SPIE 6745, Remote Sensing of Clouds and the Atmosphere XII, 67450J (October 03,2007), doi:10.1117/12.737801.

[12] P. Pešice and Z. Sokol, "Meteosat parallax correction using radar echotops," ERAD 2008 – The fifth European Conference on radar in Meteorology and Hydrology. ISBN 978-951-697-676-4, CD.

[13] C. Morel and S. Senesi, "A climatology of mesoscale convective systems over Europe using satellite infrared imagery. I: Methodology," Quarterly J. of Royal Meteorological Society, 2002, vol. 128, pp. 1953-1971.

[14] S. Kolios and H. Feidas, "A warm season climatology of mesoscale convective systems in the Mediterranean basin using satellite data," Theoretical and Applied Climatology, 2010, vol. 102, pp. 29-42.

[15] D. J. Seidel, B. Sun, M. Pettey, and A. Reale, "Global radiosonde balloon drift statistics," J. of Geophysical Research, 2011, doi:10.1029/2010JD014891.

[16] J. Wang and W. B. Rossow, "Determination of cloud vertical structure from upper-air observations," Journal of Applied Meteorology. 1995, vol. 34, pp. 2243-2257.

[17] P. K. Simpson, "Artificial neural system-foundation, paradigm, application and implementation," New York: Pergamon Press, 1995.

[18] C. Bishop, "Neural Networks for Pattern Recognition," Oxford, University Press, 1995.

[19] S. Weisberg, "Applied Linear Regression," Willey, 2005.

[20] V. Vapnik, "The Nature of Statistical Learning Theory," Springer, New York, 1995.

[21] A. Smola and B. Schölkopf, "A tutorial on support vector regression", Journal Statistics and Computing, 2004, vol. 14(3), pp. 199-222.

[22] C-W. Hsu, C-C. Chang, and C-J. Lin, "A practical guide to support vector classification," Technical report, Department of Computer Science, National Taiwan University. July, 2003.

[23] L. Breiman, "Random Forests, Machine Learning," 2001, vol. 45(1), pp. 5-32.

 ISBN: 978-1-61208-264-6