IEEE-INNS IJCNN - International Joint Conference on Neural Networks, Budapest, 2004.

Hybrid Neural-Phenomenological Sub-Models and its Application to Earth-space Path Signal Attenuation Prediction

Luiz P. Calôba
Electric Eng. Progr.
COPPE & EP/UFRJ
Rio de Janeiro, Brazil
caloba@ufrj.br

Gilson A. Alencar
Electric Eng. Dept.
UGF
Rio de Janeiro, Brazil
gilsonalencar@uol.com

Mauro S. Assis
Electric Eng. Dept.
IME
Rio de Janeiro, Brazil
mauroassis@openlink.com.br

Abstract - Neural models may be very precise but, being numerical, provide only limited contribution to the understanding of the phenomenological process, contrary to phenomenological models. In this paper we use a neural techniques to evaluate and to provide information on the sub-models that composes a phenomenological model. We also show how some hybrid neural-phenomenological sub-models may be used to maximally preserve the phenomenological information while providing numerical precision.

The problem of radiowave degradation by rain is critical for the design of reliable earth-satellite communication links operating above 10 GHz. Phenomenological models available in the literature are complex and show poor accuracy, and so are good candidates for the proposed technique. The use this technique in the UIT-R model presented very interesting results.

I. NEURAL INTERPRETATION OF PHENOMENOLOGICAL MODELS.

Mathematical models usually consists of a set of non-recursive equations successively applied, each one generating an internal variable of the model. Generally each equation has a phenomenological meaning, and may be considered as a phenomenological sub-model integrating the model.

Representing each equation as a block, or sub-model, we have the block diagram of the global model. The fig 1 presents a block seen as a sub-model, and the fig 2 presents the block diagram of the global model, constituted by several sub-models.

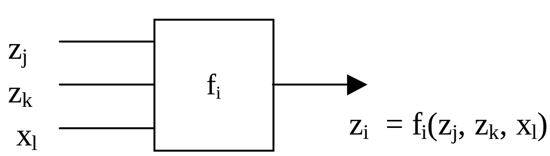

Fig. 1 - Block or sub-model

Representing the connections between blocks in fig 2 by directed branches, we have a kind of feedforward signal flow graph. Interpreting the branches as unitary gain synapses and each block as a neuron with excitation/activation function equal to the equation it represents, f_i, clearly the phenomenological model corresponds to a non-conventional feedforward neural network. Models or sub-models using recursive equations will not be considered in this paper but, analogously, they may be represented by non-conventional recursive neural networks, e.g. TDNNs.

II. BLOCK ERROR

Unless its equation represents with sufficient precision the phenomenological sub-model, each block introduces an error that composes the total error at the model output. It is important to classify the blocs in function of their contribution to the error at the model output. Consider a synapse c_i , added to the output of the block f_i to correct the block's error, as shown in fig 3.

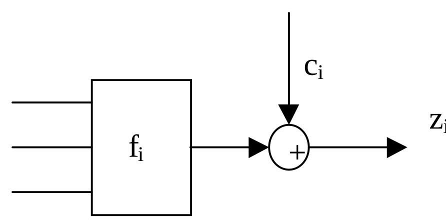

Fig 3 - Block with error correcting synapse

If the block diagram in fig. 2 perfectly represents the phenomenological model, the nominal value of c_i must be zero for each block *i* and each input-output pair. Our objective now is to measure the contribution of each block, or sub-model, to the output error. Consider the objective function F(.) ,

$$F = \underset{\forall n}{E}(e_n) = \frac{1}{N}\sum_{n=1}^{N} e_n \qquad (1)$$

$$e_n = e\left(y_n, \tilde{y}_n\right) \qquad (2)$$

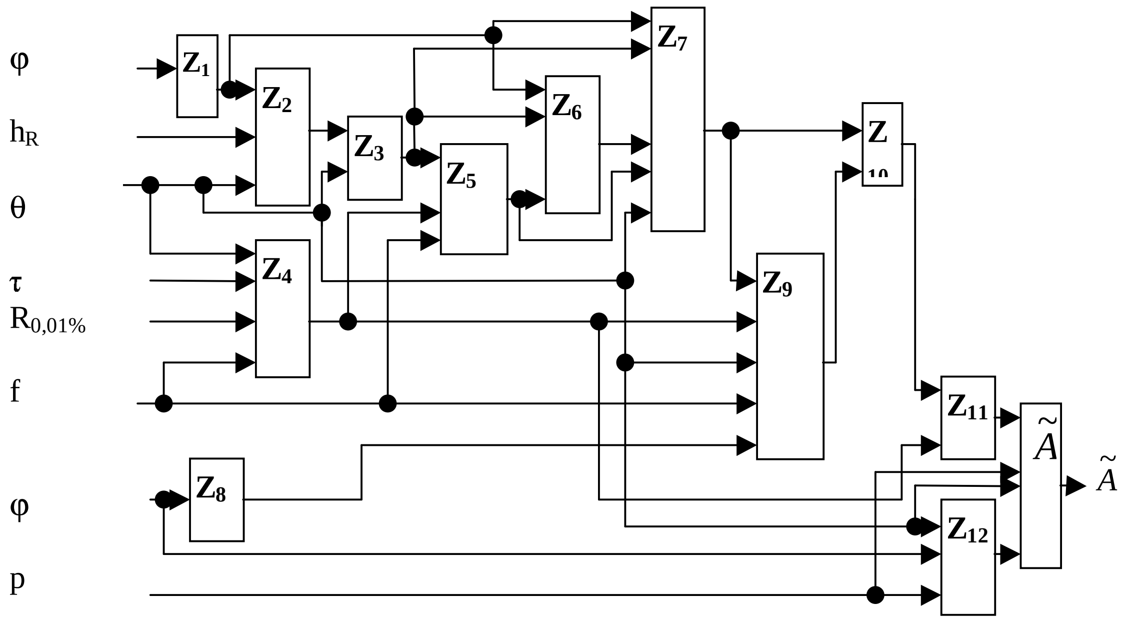

Fig 2 - Block diagram of the UIT-R model

where E(.) is the expected value, for all input-output pair n = 1,...,N, e(.) is the output error, e_n is the value of e(.) for the input-output pair n, y_n is the target value and $\tilde{y}_n$ the actual model output value for pair n.

Considering the model as a non-conventional neural network, c_i may be adjusted by the descendent gradient of F to a vector $\underline{c}$ with components c_i . So,

$$\Delta c_i(n) = -\alpha \frac{\partial e_n}{\partial c_i} \qquad (3)$$

$$\frac{\partial e_n}{\partial c_i} = -\alpha \frac{\partial e_n}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial z_i} \frac{\partial z_i}{\partial c_i} \qquad (4)$$

$$\Delta c_i(n) = -\alpha \frac{\partial e_n}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial z_i} = -\alpha \delta_i \qquad (5)$$

where $\alpha > 0$, $\partial z_i / \partial c_i = 1$ and δ_i is the output error backpropagated to the output of the block *i* for pair n. As it is a descendent gradient error optimization, for a first order approximation of F and for constant $|\underline{\Delta c}|$ the calculated value of $\underline{\Delta c}$ minimizes F. On the other hand, when $\Delta c_i(p)$ completely compensates the error introduced by the block *i* , the output error vanishes, and so does F. So, δ_i is a very adequate parameter to compare the importance of the error generated by each block in the output model error.

The classical synapses correction when training a neural network is

$$\overline{\Delta c_i} = E_{\forall p}\left[\Delta c_i(p)\right] \qquad (6)$$

This correction is implemented by a bias at the block output. However, this correction is very poor, correcting only the mean value of the error, which is already very small if the blocks were properly projected. It is more effective to correct the error $\Delta c_i(p)$, different for each pair p. And so, a good measure of the influence of each block error in the output error is the variance of $\delta_i(p)$, eq. 7.

$$\sigma_i^{\,2} = E_{\forall p}\left[\delta_i^{\,2}(p)\right] - E_{\forall p}^2\left[\delta_i(p)\right] \qquad (7)$$

The critical blocks are those with higher influence on the output error, i.e., those with higher $\sigma_i^{\,2}$. Those blocks need to be fixed to provide a more precise global model.

III. Analysis and Correction of Sub-Models.

A sub-model may be inadequate by two main reasons: its equation is inadequate and / or its input variables do not provide sufficient information to correctly determine its output variable value. If the block error (i.e. the general model output error backpropagated to the block output) presents significant correlation with any of the block input variables, probably the sub-model equation is not adequate to represent it, and must be fixed. If this correlation does not exists, than probably there is a lack of information at the block inputs. When a block equation is calculated, the general model inputs and some of its internal variables (other blocks outputs) are already available. The information needed at the block input may be present in those available variables. If some of those variables present significant correlation with the block error, that may be used as an additional block input to reduce its error.

Usually both reasons occur simultaneously. In this case the available variables that presents correlation with the block error must be de-correlated from the block input variables that also presents that correlation. If the de-correlated additional variables does not maintain correlation with the block error, they must be discarded. But if the correlation still exist, the additional variables, de-correlated or not, must be considered as possible additional inputs to the block. This last step, involving the de-correlation, avoids the introduction of unnecessary variables at the block input and is essential to maximally preserve the phenomenological sub-models.

The above procedure allows determining the critical blocks, that need to be fixed, and the additional variables that may be used as new inputs. The correction is made by changing the sub-model equation, which is not an easy task, mainly if the blocks have been carefully designed, as usual. An alternative is to use an hybrid neural-phenomenological sub-model: a neural network is placed in parallel (or in cascade) with each critical block, as shown in fig. 4. The goal of this network is to generate the symmetrical value of the block error, vanishing it.

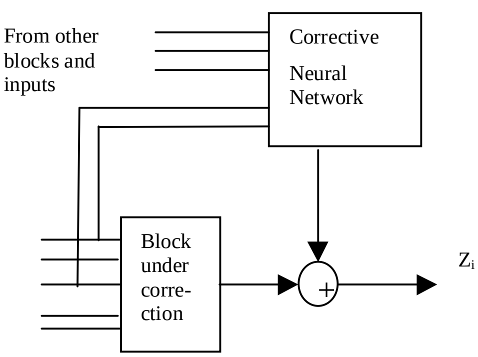

Fig 4 - Block error correction using a local neural network.

The neural network is fed by the block inputs that present correlation with the block error, and / or by the additional available variables that present correlation with the block error even when de-correlated from the block input variables, as discussed above.

Those auxiliary neural networks must be trained simultaneously, embedded in the phenomenological model. If the training procedure is based on the backpropagated error. The error backpropagated from the general model output $\tilde{y}$ to the block output z_i is

$$\delta_i = \frac{\partial e_p}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial z_i} \qquad (7)$$

From this point on the error backpropagation inside the neural network is as usual. As the inputs and the internal variables of the phenomenological model are not normalised, it is recommended to normalise it to zero mean and (-1,+1) range before presenting it to the neural network, to avoid ill-conditioned training. The same should be done with the block error, the goal of the neural network. However, if the output neuron of the network is linear, this last normalisation is not strictly necessary.

IV. Earth-Satellite Communication Path Models

The satellite communication services have grown on last years and the radiowave spectrum to support them is saturated. So, it’s necessary to search for frequency bands higher than the presently used ones, to allocate new services. But the problem of radiowave degradation by rain is critical for communication links operating above 10 GHz, and a precise knowledge of rain attenuation is important to design reliable satellite communication links, considering that it must operate under all atmospheric conditions. The attenuation A(p) that is not exceeded in a time percentage p of the average year is a critical parameter in the design of those links. Several phenomenological models (e.g. UIT-R - or ITU-R, the Radio-communication Sector of International Telecommunication Union [2], American [3], Japanese [4], Spanish [5] and Brazilian [6]) have been developed to predict the rain attenuation in earth-space paths, but these models show poor accuracy for higher frequencies. In this paper we choose to present results on the UIT-R model, because in the literature it is used as reference and comparison with all other models.

In this paper we used the data bank from UIT-R [1], relating frequency (f), polarity angle (τ), elevation angle (θ), latitude (φ), rain height (h_R) and rain rate ($R_{0.01\%}$) that is exceeded for 0.01% of the average year, to the attenuation {A(p)} that is not exceeded for a time percentage (p) of the average year. We considered time percentages from p = 0.001 to 0.5% and frequencies from f = 11 to 20 GHz. Unfortunately only a small number, 80 to 165, of input-output pairs are available, depending on the time percentage. To hold with this problem a selection criteria of the input-output pairs was developed in order to guarantee a reasonable generality for the training (70% of the pairs) and test (30% of the pairs) sets [7]. Moreover, some critical variables in the data bank, e.g. rain rate and attenuation, are not very precise. But this is the only data bank available in the world.

The root mean-squared (RMS) relative error, E, eq. (8), is the demerit factor proposed by UIT-R to evaluate the performance of a model.

$$E(\underline{w}) = \sqrt{F(\underline{w})} \qquad (8)$$

$$F(\underline{w}) = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{A_n - \tilde{A}_n}{A_n}\right]^2 \qquad (9)$$

where $\underline{w}$ is the synapses vector, A_n is the measured attenuation and $\tilde{A}_n$ is the attenuation estimated by the model for the input-output pair n. N is the total number of a input-output pairs in the training set.

To reach the best performance, the objective function used to train the neural network must minimise the demerit function proposed by UIT-R, the mean-squared relative error E shown in eq. 8, or the function F in eq. 9. The input-output pairs are constituted by the seven inputs, f, h_R, $R_{0.01\%}$, , τ, φ, θ , p, and the desired output, A(p).

V. The UIT-R Model

Analysing the UIT-R model we verify that it is composed by 12 internal sub-models, whose outputs are Z_1 to Z_{12}, connected as shown in fig. 2. Each sub-model has a phenomenological meaning; its equations are:

$Z_1 = f(\varphi) = h_R$

$$\begin{aligned} h_R &= 5 - .075(\varphi - 23) && \text{for } 0 \le \varphi \le 23 \\ h_R &= 5 && \text{for } -21 \le \varphi \le 0 \\ h_R &= 5 + .1(\varphi + 21) && \text{for } -71 \le \varphi < -21 \\ h_R &= 0 && \text{for } \varphi < -71 \end{aligned}$$

$Z_2 = f(Z_1, h_R, \theta) = L_S$,

$$L_S = \frac{(h_R - h_S)}{\operatorname{sen}\theta} \qquad \text{for} \quad \theta \ge 5^0$$

$$L_S = \frac{2(h_R - h_S)}{\left(\operatorname{sen}^2\theta + \frac{2(h_R - h_S)}{R_e}\right)^{1/2} + \operatorname{sen}\theta} \qquad \text{for } \theta < 5^0$$

$Z_3 = f(Z_2, \theta) = L_G = L_S \cos\theta$

$Z_4 = f(f, \tau, \theta, R_{0,01\%}) = \gamma_R = k\left(R_{0.01\%}\right)^{\alpha}$

$Z_5 = f(Z_3, Z_4, f) =$

$$r_{0.01} = \frac{1}{1 + 0.78\sqrt{\frac{L_G\gamma_R}{f}} - 0.38\left(1 - e^{-2L_G}\right)}$$

$$Z_6 = f(Z_1, Z_3, Z_5) = \xi = \tan^{-1}\left(\frac{h_R - h_S}{L_G r_{0.01}}\right)$$

$Z_7 = f(Z_1, Z_3, Z_5, Z_6, \theta) = L_R$

$$\text{for } \zeta > \theta,\quad L_R = \frac{L_G r_{0.01}}{\cos\theta}$$

$$\text{else,}\quad L_R = \frac{(h_R - h_S)}{\operatorname{sen}\theta}$$

$Z_8 = f(\varphi) = \chi$

$$\text{if } |\varphi| < 36^0,\ \chi = 36 - |\varphi|$$

$$\text{else,}\quad \chi = 0$$

$Z_9 = f(Z_4, Z_7, Z_8, f, \theta) =$

$$v_{0.01} = \frac{1}{1 + \sqrt{\operatorname{sen}\theta}\left(31\left(1 - e^{-(\theta/(1+\chi))}\right)\frac{\sqrt{L_R\gamma_R}}{f^2} - 0.45\right)}$$

$$Z_{10} = f(Z_7, Z_9) = L_E = L_R v_{0.01}$$

$$Z_{11} = f(Z_4, Z_{10}) = A_{0.01} = \gamma_R L_E$$

$Z_{12} = f(\theta, \varphi, p) = \beta$

if $p \ge 1\%$ or $|\varphi| \ge 36^0$, $\beta = 0$

if $p < 1\%$ and $|\varphi| < 36^0$ and $\theta \ge 25^0$,

$$\beta = -0.005\left(|\varphi| - 36\right)$$

$$\text{else,}\quad \beta = -0.005\left(|\varphi| - 36\right) + 1.8 - 4.25\operatorname{sen}(\theta)$$

$$\tilde{A}_p = f(z_{11}, z_{12}, p, \theta) =$$

$$= A_{0.01}\left(\frac{p}{0.01}\right)^{-(0.655 + 0.033\ln(p) - 0.045\ln(A_{0.01}) - \beta(1-p)\operatorname{sen}(\theta))}$$

The parameters k and α in Z_4 equation are not computed by blocks of the model, but from the Marshall and Palmer distribution, and so are considered error free.

VI. Sub-Models Precision and Critical Sub-Models

The error backpropagated from the model output to the z_i sub-model output for pair n is

$$\delta_i(n) = \frac{\partial}{\partial \tilde{A}_n}\left[\frac{A_n - \tilde{A}_n}{A_n}\right]^2 \frac{\partial \tilde{A}_n}{\partial z_i} = -2\left(\frac{A_n - \tilde{A}_n}{A_n^2}\right)\frac{\partial \tilde{A}_n}{\partial z_i} \qquad (10)$$

The derivative $\partial \tilde{A}_n / \partial z_i$ was calculated numerically, introducing a small, $\pm 10^{-6}$, increment in Z_i , and taking care with some discontinuities in some block equations. The standard deviation of δ_i was calculated for each block and for each time percentage, and the results are shown in Table 1

Table 1
Standard deviation of δ_I for each sub-model and for the different time percentages.

	1	2	3	4	5	6	7	8	9	10	11	12	$\tilde{A}$
0.001	0.21	0.09	0.10	0.78	**1.25**	0.00	0.15	0.02	**1.16**	0.19	0.15	1.17	0.36
0.002	0.17	0.08	0.09	0.68	**0.96**	0.00	0.13	0.01	**0.93**	0.16	0.14	0.64	0.36
0.003	0.17	0.07	0.08	0.63	**0.82**	0.00	0.11	0.00	**0.81**	0.15	0.13	0.43	0.30
0.005	0.15	0.07	0.08	0.58	**0.74**	0.00	0.10	0.01	**0.72**	0.13	0.12	0.21	0.27
0.01	0.13	0.06	0.08	0.49	**0.85**	0.00	0.11	0.00	**0.78**	0.14	0.11	0.00	0.29
0.02	0.15	0.07	0.09	0.46	**1.10**	0.00	0.13	0.00	**0.94**	0.17	0.11	0.26	0.31
0.03	0.17	0.08	0.10	0.46	**1.30**	0.00	0.15	0.00	**1.08**	0.19	0.12	0.44	0.32
0.05	0.19	0.09	0.12	0.48	**1.72**	0.00	0.18	0.00	**1.37**	0.15	0.15	0.76	0.35

Table 1 evidences that sub-models 5 and 9 are the critical ones, followed by sub-models 12 and 4. We decided to fix only blocks 5 and 9, in order not to change too much the phenomenological structure.

Those results agree with physical analysis. The UIT-R model considers uniform rain in the path, which is evidently not true. To fix this approximation, heuristic formulas are used as "horizontal reduction factor", $r_{0.01}$, and "vertical adjust factor", $v_{0.01}$, established for p = 0.01%. Those heuristic factors are calculated by sub-models 5 and 9.

VII. Fixing the Critical Sub-Models

The first step to fix the sub-model *i* is to verify the correlation of δ_i with the available variables. The correlation was initially calculated for p = 0.01%, where a larger (165) number of pairs were available. The results are presented in Table 2.

Table 2
Backpropagated error correlations for sub-models 5 and 9, determined for p=0.01%.

Variables	Block 5	Block 9
Frequency, f	0.01	0.01
Polarization, τ	0.00	0.06
Elevation, θ	-0.11	-0.05
Latitude, φ	**-0.30**	**-0.35**
Height, h_R	**-0.25**	**-0.24**
Rain, $R_{0.01\%}$	0.13	0.15
Z_1	0.10	0.13
Z_2	**0.21**	**0.18**
Z_3	**0.20**	**0.16**
Z_4	0.12	0.13
Z_5		**-0.21**
Z_6		0.02
Z_7		0.12
Z_8	**0.22**	**0.23**
Z_9		-0.09
Z_{10}		
Z_{11}		
Z_{12}	**0.28**	**0.24**

The total number of pairs for p = 0.01% was N = 165 and so the 95% confidence level is $2\sigma_r = 2/\sqrt{N} = 0{,}156$. The bolded numbers represents significant correlations. The variables that presented significant correlation with the block errors were then de-correlated form the block input variables (and also for other variables already accepted as inputs for the corrective neural network), and the correlation with the block error was tested again. Those that preserved the correlation with the block error after de-correlation were then selected as inputs of the corrective neural network. A similar procedure was applied to the other time percentages, and in consequence other variables were introduced. The final selection of variables to present at the input of the corrective neural networks is shown in Table 3.

Table 3
Selected variables for the inputs of the corrective neural networks for blocks 5 and 9.

Block 5	Block 9
Latitude, φ	Latitude, φ
Height, h_R	Height, h_R
Elevation, θ	
Z_2	Z_2
Z_3	Z_3
	Z_7
	Z_8

A neural network was then connected in parallel with each critical block, 5 and 9, as shown in fig 3. Some preliminary experiments had shown that 10 neurons with hyperbolic tangent as activation function at the intermediate layer and one linear neuron at the output layer was an adequate choice for our problem. The neural networks undergo a supervised training that is a modified version of the well-known back-propagation algorithm, with overtraining control. Both neural networks were trained simultaneously, taking the cares discussed in earlier sections.

VIII. Results and Models Comparison

For the sake of comparison we also developed a neural network to implement the whole attenuation model [7]. As this is a more complex mapping problem, this network required 15 neurons in the intermediate layer. The results for the three models, the pure phenomenological UIT-R Model, the hybrid phenomenological-neural, UIT-R-Neural Model, developed in this paper, and the pure Neural Model [7] are presented in Tables 4 and 5.

TABLE 4
RMS RELATIVE ERROR (%) FOR ALL PAIRS.

Model / Time Percentage	UIT-R	Hybrid UIT-R - Neural	Neural
0,001%	38	21	17
0,002%	32	20	20
0,003%	30	23	20
0,005%	28	19	18
0,01%	29	21	20
0,02%	31	24	21
0,03%	32	22	22
0,05%	35	23	20

TABLE 5
RMS RELATIVE ERROR (%) FOR THE TEST SET ONLY.

Model / Time Percentage	UIT-R	Hybrid UIT-R - Neural	Neural
0,001%	25	22	22
0,002%	23	17	19
0,003%	28	24	19
0,005%	21	17	20
0,01%	28	23	22
0,02%	27	24	22
0,03%	27	22	24
0,05%	26	23	22

As expected, the best results were obtained using the Neural Model, that is an universal approximator. The second best results, very close to those of the Neural Model, were those obtained using the Hybrid Model, which has the advantage of preserving the phenomenological information. The worst results were those from the phenomenological model, but the earth-space rain attenuation is a known as a hard to model problem. The best performance of the phenomenological model in the test set is because its adjustment uses all pairs.

IX. Conclusions

Neural techniques may furnish valuable information on phenomenological models, determining its critical sub-models and giving insights on how to fix them. If it is not possible to fix them using phenomenological knowledge, one alternative is to use a local neural network to fix the output of the critical sub-models. This hybrid phenomenological-neural model furnishes the precision of the numerical neural model while preserving the phenomenological model information. This technique was applied to the prediction of the rain attenuation on earth-satellite communication paths with very interesting results.

Acknowledgment

The authors acknowledge the financial support of CNPq, CAPES and FAPERJ to this work.

References

[1] UIT-R, "Acquisition Presentation and Analysis of Data in Studies of Tropospheric Propagation", Recommendation UIT-R PN.311-6, Geneva, 1994.
[2] UIT-R, 2000, "Rain Attenuation Prediction Methods", Recommendation 618-4, Geneva, 2000.
[3] CCIR (Now ITU), 1993, Prediction of rain attenuation-proposed amendment to Recommendation 618, *Doc. 5C/56*, Geneva, 1993.
[4] CCIR (Now ITU), 1993, Proposal of an improved rain attenuation prediction method, *Doc. 5C/69*, Geneva, 1993.
[5] CCIR (Now ITU), 1993, Modifications to the microwave rain attenuation calculation method, *Doc. 5C/81*, Geneva, 1993.
[6] CCIR (Now ITU), 1993, Slant-path attenuation method based rainfall rate distribution, *Doc. 5C/52*, Geneva, 1993.
[7] Alencar,G.A.; Calôba,L.P.; Assis, M.S., "Artificial Neural Networks as Rain Attenuation Predictors in Earth-Space Paths", IEEE-ISCAS, Bangkok, 2003.