Statistical Models of the Troposphere Refractive Index

Kravchenko V.F.¹, Lutsenko V.I.²*, Lutsenko I.V.², Popov D.O.²

¹ Kotel’nikov Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, 125009, Russia
²Usikov Institute of Radiophysics and Electronics NAS of Ukraine, Kharkov, 61085, Ukraine
*Corresponding Author: lutsenko@ire.kharkov.ua

Abstract The statistical model for describing the refractive index of the troposphere based on nested semi-Markov processes is proposed. The possibility of describing the statistics of the refractive index in season by locally Gaussian model is shown. First considered the use of atomic functions (AF) to describe the statistics of the refractive index, a new approach for testing hypotheses about the form of the random variable, based on the use of inverse functions is proposed.

Keywords Gaussian Model, Refractive Index, Finite Atomic Functions, Semi-Markov Processes, Kravchenko Functions

1. Introduction

The effectiveness of the navigation systems, radar, communication depends heavily the radio propagation conditions, determined by the state of atmospheric refraction. It, in its turn, is caused by the spatial and temporal distribution of the refractive index \( N \). Therefore, the study of seasonal, diurnal and vertical distribution of \( N \), making of its statistical model provides the great interest to predict the range of the radio systems for various applications, as well as estimates of the accuracy of measuring the coordinates of global navigation satellite systems (GNSS). Productive, at the same time may be the approach which was used earlier to describe the space-time non-stationary signals scattered underlying surfaces and "clear" sky using nested semi-Markov processes [1, 2].

For experimental studies of the influence propagation conditions of radio waves on the estimation error of coordinates in the cities of Kharkov, Poltava, Sumy, Smela were set up the measuring points, equipped with receiving equipment devised by Ltd. "Navis-Ukraine", where were conducted round the clock systematic measurement of errors of estimating the coordinates. In meteorological providing of measurements are used accumulated data of regular meteorological stations in these cities [3].

2. The Models of Approximation of the Experimental Refractive Indices

2.1. The Mathematical Formulation of the Model

Statistical description of the behavior of the refractive index can be based on the use of nested two-component stochastic processes \( \{S(t), \theta(t)\} \), of which one component \( S(t) \) is continuous, the second \( \theta(t) = \nu_j \) - a discrete and \( t \) time. These components are sensitive and generally not Markov [4, 5]. This means that no restrictions are imposed on the distribution of the times of a process in each of the phase states. At each moment of time the process is in one of the \( K = 4 \) possible phase states \( H_i \in \nu_1...\nu_K \) (\( \nu_1 \) - summer, \( \nu_2 \) - autumn, \( \nu_3 \) - winter, \( \nu_4 \) - spring), and the initial state is assumed known at the time \( t_0 = \nu_j \). Since the change of seasons is a deterministic process, then the one-step transition probability \( \pi = [\pi_{ij}] \), where \( i, j = 1...K \) have the following form: \( \pi_{ij} = 1 \) at \( j = (i+1)_{\mod 4} \) and \( \pi_{ij} = 0 \) at other times. For each nonzero element \( \pi_{ij} \) of the transition probability matrix is comparable to a random variable \( T_{ij} \) with density function \( f_{ij}(t) \), which will call the time-out in the state \( \nu_j \) before the transition to the state \( \nu_j \), and the only possible transitions \( \nu_j \Rightarrow \nu_j \), and besides \( j = (i+1)_{\mod 4} \). The waiting time \( T_{ij} \) will depend on the average length of the season \( \overline{T}_{ij} \). Since the values \( T_{ij} \) are not distributed exponentially (more than justified the use of the normal distribution), then this will be a semi-Markov process. Within each of the states, it is quasi-stationary with the density values \( [P_i(N)] \) of the refractive index \( N \) and its spectrum \( [S_i(\omega)] \), which takes into account the temporal behavior of the refractive index \( N(t) \) in each of the seasons.

Mathematical model of the system that describes the seasonal behavior of the refractive index are nested processes [4] one of which is a semi-Markov [5] determines the change of phase states, and the second determines the behavior of the system in phase state. In our case, the density can be described by locally Gaussian distribution [6-8] or by Kravchenko distribution [9-10].
2.2. The Use of Poly-Gaussian Models

Using the results of systematic measurements (for some cities 8 times a day, and for the other 4 times), meteorological parameters (temperature, pressure, and humidity) by regular weather stations in Ukraine were assessed values of the refractive index \([11, 12]\) and formed a database of all regions of Ukraine. In total were covered 100 cities for the period from 01.01.2010 to 01.01.2012. In the winter, regardless of where the city is located (on the sea coast or deep in the interior) it's about 310 N-units. In the summer, depending on the location of the city on the sea coast or in the interior of the country it is from 380 N-units to 350 N-units. Significant non-stationary of daily and seasonal behavior is manifested in non-Gaussian of its distribution, the most notable at the values of the refractive index substantially larger or smaller than the average value. The investigation of distribution inside of each of the seasons, showed that it is in the first approximation satisfactorily described by the standard Gaussian model \([6]\). The final probabilities of each of the phase states \(P_i\) were determined and shown that it allows approximately 70% of cases significantly reduce the approximation error compared to using a uniform probability model \(P_i = 0.25\) for the seasons \(i = 1...4\). For the cities in which we measured, was examined the possibility of using to describe the density distribution of poly-Gaussian models (Fig. 1, 2).

Similarly, in Kharkov and Smela experimentally obtained distribution density for each of the seasons, as well as their approximation Gaussian probability density with the same numerical characteristics, as in the experiment (average and rms values) are shown in Fig. 2

Fig. 1 shows the density distribution of the refractive index of the troposphere for Poltava in different seasons, as well as its approximation by the Gaussian functions.

![Figure 1](image)

**Figure 1.** The Density Distribution of the Refractive Index of Poltava in Different Seasons, where \(N^* = (N - \overline{N})/\sigma_N\): a) spring, b) summer, c) autumn and d) winter (1 – Experiment 2 – approximation of the Gaussian functions)

Fig. 1 shows the density distribution of the refractive index of the troposphere for Poltava in different seasons, as well as its approximation by the Gaussian functions.

2.3. Approximation of the Density Distribution of the Refractive Index by Kravchenko Functions

In contrast to the mathematical models that is used to describe the density distribution of the refractive index of poly-Gaussian representation, the real value of its is finite. It is a prerequisite for the application of atomic functions (AF) \(up(x)\) \([10,13-15]\). It is finite differentiable function which having a bell-shaped form such that its derivative can be represented, in turn, of the two bell-functions, each of which...
is shifted and compressed copy original function up to a scale factor [10, 13]. A generalization of \( u\) is the Kravchenko AF [10, 14, 15].

Currently AF is widely used in approximation theory, numerical analysis, digital signal processing, wavelet analysis, and other areas. Functions \( u(x) \) is an even. It increases in the area \([-1, 0]\), decreases in the interval \([0, 1]\) and limited to a single area. Moreover, \( u(p-1-x) = 1-u(p)(x) \) at \( x \in [0,1] \).

Good approximation properties of the functions \( u(x) \) based on the fact that with the help of a linear combination of shifts-contractions \( u(x) \) can be any algebraic polynomial of any degree [10]. It is of interest to clarify the possibility of the use the unique properties of Kravchenko functions [9, 10, 14, 15], to describe the density distribution of the refractive index inside each of the seasons.

The function \( u(x) \) can be considered as the convolution of an infinite number of rectangular pulses and can be written in terms of the Fourier transform of the characteristic function, which in this case can be written as the product of the characteristic functions of a square pulses \( p(\xi) \) [9, 10]:

\[
up(\xi) = \frac{1}{2\pi} \prod_{j=1}^{\infty} \sin c \left( \frac{t}{2} \right) \exp \int_{-\infty}^{\infty} \xi dt \quad (1)
\]

According to (1) were calculated the values of the Kravchenko functions. As the input data used for the approximation of the numerical characteristics (mean and rms value) of the refractive index obtained from the experimental data for each of the seasons. For example, Fig. 3 shows the density distribution of the refractive index for Poltava during the year, as well as its approximation by Kravchenko functions with the same mean and rms. The variance of the approximation error in this case has a value of \( 1.08 \cdot 10^{-5} \).

The distribution density of the refractive index of the troposphere for each season from 01.03.2010 to 01.03.2011 and from 01.03.2011 to 01.03.2012 and for the year as a whole, as well as their approximating by Kravchenko functions in Kharkov and Smela are presented in Fig. 4.

Final probabilities for each of the seasons were taken the same - 0.25.
3. Checking the Assumptions on Distributions

Statistical testing of the assumptions about the distribution enables to "active" assess whether provides to the adopted model a fairly accurate description of the observed data. Typically, a standard test involves the following main steps [16]:
- Based on the obtained data computed some number, which called the goodness of fit.
- Determined the probability of obtaining the computed criteria, provided that the model was chosen correctly. Often this is done by referring to a table of percentiles of distribution criterion.
- If the probability of obtaining a calculated value of the criterion is small, then it is believed that the adopted statistical model does not give a correct description of the data. Usually considered to be "small" probability, starting with 0.10, 0.05 or even less. If the calculated probability is not "small", the data give no reason to believe that the adopted model is not suitable.

It should be noted that although this technique allows us to discard the model as wrong, but it doesn't allows us to prove that the model is correct. Verification result is largely dependent on the amount of data available: more data, gives more chance to reject the wrong model. If the data is not enough, it is often impossible to establish the inadequacy of even a model that is essentially different from that adopted.

Developed many of criteria to assess the validity of the assumptions of the distributions, some criteria are valid only for certain models, some are applicable to a wide range of distributions. Many techniques require accurate knowledge of each parameter of the statistical model. One of the most frequently used criterion to verify the assumptions of the law of distribution is the $\chi^2$-square test which can be used to test any distribution, including normal and exponential [16].

Checking assumptions using the $\chi^2$-square test is based on the calculation of the functional

$$
\chi^2 = \sum_{i=1}^{k} \frac{(M_i - E_i)^2}{E_i},
$$

where $M_i$ is a number of observations in each of $i$ interval, $E_i$ is the expected mean of the number of observations in each of the intervals according to the model, which is analyzed, $n$ is sample size, $k$ is the number of differential pockets on which distributes the data. Then

$$
\chi^2 = n \sum_{i=1}^{k} \frac{(p_i - \hat{p}_i)^2}{p_i},
$$

where $p_i$ is the experimental density of distribution, $\hat{p}_i$ is the density distribution according to the model.

Thus, on this criterion determined discrepancy between the theoretical model and experimental data which are normalized to the theoretical density for this pocket. In this approach is desirable to choose the width of the pocket in that way that the probability of hitting were almost identical. However, this is not always possible to implement. In many cases, the same pockets are used. At the same time employed criterion for the low levels of probability weakly takes into account the smallness of their value. It gives special trouble in the solution of problems of radiolocation, when in calculations of the performance characteristics of systems interested in small values of probability of false alarm (typical values less than $10^{-4}$). No account of differences of probability distributions from those which used to calculate the models at low levels of probability leads to significant errors in the evaluation of performance characteristics of systems. Such criteria for testing hypotheses about the form of the experimental distribution law that allows equally well to assess differences between experimentally obtained distributions and theoretical models for large and small significance levels are required.


Suppose that there is some experimentally obtained histogram of the distribution of the refractive index and the integral distribution function. It is necessary to check several hypotheses:

4.1. The Possibility of Approximating the Experimental Data by the Set of Several Normal Distributions

Poly-Gaussian representation of a random variable:

$$
\hat{p}(N) = \sum_{i=1}^{k} a_i \phi_i(N) \quad \tilde{F}(N) = \sum_{i=1}^{k} a_i \Phi_i(N),
$$

where $\phi_i(N) = \frac{1}{\sqrt{2\pi} \sigma_{N_i}} \exp \left( \frac{(N - m_{N_i})^2}{\sigma_{N_i}^2} \right)$ is the locally normal distribution density of the refractive index for $i$ phase state (the season) with the average value $m_{N_i}$ and RMS $\sigma_{N_i}$, $\Phi_i(N)$ its integral distribution:

$$
\Phi_i(N) = \int_{0}^{N} \phi_i(N)dN, \quad a_i \quad \text{are the approximation coefficients}
$$

which are determined by the method of least squares and provide minimization of norm of the difference between the experimentally obtained density distribution $p(N)$ and the approximating function $\hat{p}(N) : \frac{\partial}{\partial a_i} \left\| p(N) - \hat{p}(N) \right\| = 0$.

4.2. Approximation of the Refractive Index of the Troposphere by Kravchenko Functions

This is a fairly broad class of functions. To describe the
distribution density and integral distributions are applied Kravchenko functions, denoted like \( kr_i(N) \) for the distribution density and \( KR_i(N) \) for the integral functions. Approximation also searched by the same method of least squares and finds the optimal set of coefficients \( a_i \)

\[
\hat{p}(N) = \sum_{i=1}^{k} a_i kr_i(N), \quad \hat{F}(N) = \sum_{i=1}^{k} a_i KR_i(N). \tag{5}
\]

To test the hypothesis how well an approximation based on a model, describes the experimental data, is checked the value of difference (norm) between the integral function of the experimental data, which transformed by the inverse function of approximating model and a linear function. If \( \hat{F}(N) = y \) is approximation of the integral distribution function, for example, of the refractive index \( N \), to \( F^{-1}(y) \) is inverse function. Obviously, when the exact correspondence of the approximating model with experimental then the points lie on a straight line \( y = N \), since

\[
F^{-1}(y) = \hat{F}^{-1}(F(N)) = N. \tag{6}
\]

By constructing the dependence \( y = \hat{F}^{-1}(F(N)) \) and evaluating its deviation from a straight line \( y = N \) can be judge about difference between obtained from the experimental distribution function and used for its approximation. This tests the hypothesis of the possibility of describing by the approximating model which used \( \hat{F}(N) \). Comparison is made in the space formed by mapping the experimental distribution \( F(N) \) function by the inverse of the approximating function \( \hat{F}^{-1}(y) \).

\[
\delta^2 = \left\| N - \hat{F}^{-1}(F(N)) \right\|_2^2 = \int_{0}^{\infty} \left( N - \hat{F}^{-1}(F(N)) \right) dN. \tag{7}
\]

For the example, consider the use of the proposed approach for the approximation of the experimental data of the refractive indices of the troposphere. Use the following models:

1) poly-Gaussian approximation,
2) approximation by the set of Kravchenko functions.

### 4.3. Comparison of Distribution Models with Experimental Data

For comparison of the distribution models and the experimental data it is necessary to calculate the real and approximate distribution functions of the refractive index. For each probability from 0.01 to 0.95 in steps of 0.01 can be obtained by appropriate values of the refractive index \( N \) (real), \( N^1 \) (for the Gaussian functions) and \( N^2 \) (for the Kravchenko functions). Fig. 5 shows the dependence of \( N \) on \( N^* \) for Poltava and Smela for the period from 01.03.2010 to 01.03.2011.

Obtained with approximation by the method of least squares of the variance of the approximation error by Gaussian functions and Kravchenko functions are shown in Table 1.

<table>
<thead>
<tr>
<th>City</th>
<th>Year</th>
<th>Gauss</th>
<th>Kravchenko</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poltava</td>
<td>2011</td>
<td>9.22e-6</td>
<td>6.28e-5</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>9.85e-6</td>
<td>7.69e-5</td>
</tr>
<tr>
<td>Smela</td>
<td>2011</td>
<td>2.89e-5</td>
<td>9.62e-4</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>2.96e-5</td>
<td>8.43e-4</td>
</tr>
<tr>
<td>Cherkassy</td>
<td>2011</td>
<td>2.89e-5</td>
<td>9.19e-4</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>3.07e-5</td>
<td>2.94e-4</td>
</tr>
<tr>
<td>Sumy</td>
<td>2011</td>
<td>2.87e-5</td>
<td>4.43e-4</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>3.02e-5</td>
<td>1.38e-4</td>
</tr>
<tr>
<td>Kharkov</td>
<td>2011</td>
<td>2.96e-5</td>
<td>9.72e-4</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>3.07e-5</td>
<td>0.44e-4</td>
</tr>
</tbody>
</table>

Final probabilities for each of the seasons were taken the same 0.25.

Analysis of the data shows that for the approximation of the density distribution of the refractive index of the troposphere can be used Gaussian and Kravchenko functions, but the variance of the approximation error, which is obtained when using a poly-Gaussian model is somewhat lower than when using the Kravchenko functions, although in summer for the approximation the better results for the can provide Kravchenko functions.
The same conclusion can be arrived by analyzing the applicability of each of these hypotheses using our proposed approach. Qualitative picture of the comparison is shown in Fig. 5. Quantitative estimates of the variance of the deviations of the experimental data from linear relationship with use like inverse function the Gaussian approximation or Kravchenko functions are shown in Table 2.

As can be seen, the approximation which using Gaussian functions and Kravchenko functions gives discrepancy with the experiment on the "tails" of the distribution (Fig. 5). The greatest differences of experimental data from the Gaussian model are observed in the summer and in the transitional seasons (spring and autumn). At the same time, the approximation by Kravchenko functions largest differences are also observed in the transitional seasons (spring and fall) and winter. In case if built according to the proposed method of "inversing" of the distribution function curve lies at large values (in this case, the refractive index) above the line, then the experimental distribution function has a higher level of the "tail" of the distribution, than the model. If it is lower, then the experimental data have lower level than the model. From the relationships which shown in Fig. 5 can be seen that the Kravchenko functions are in almost all cases (except in summer) have the level of the "tails" of the distribution are lower than the experimentally obtained. This means that for a given level of significance of the refractive index obtained from the evaluation of its model will be somewhat underestimated relative to the actual values. At the same time the Gaussian functions have substantially for all seasons tails higher than the actual data. This means that they will produce excessive in relation to the actual values.

To compare the quality of the approximation obtained by using the various models can be used comparing of variance of the bias of the experimental distributions which are converted by inverted function of tested model from linearity with using Fisher's exact test \( F \) [17]

\[
F = \frac{\sigma_i^2}{\sigma_j^2},
\]

where \( \sigma_i^2 \) is the variance of error in using the \( i \) tested hypothesis of the form of distribution.

Using the relations (8) and data in Table 2 can be seen that the Gaussian functions gives better results in the approximation of the refractive index in autumn. At the same time, Kravchenko functions in winter, spring and summer somewhat better describe the behaviour of it distribution function.

### Table 2. Dispersion deviation of the distribution functions of the experimental data, transformed by using the inverse function of distribution of the tested hypothesis from a linear dependence

<table>
<thead>
<tr>
<th>City</th>
<th>Year</th>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
<th>Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smela</td>
<td>2010</td>
<td>0.1223</td>
<td>0.1073</td>
<td>0.1831</td>
<td>0.0609</td>
<td>0.1404</td>
<td>0.0913</td>
<td>0.1156</td>
<td>0.1155</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>0.0873</td>
<td>0.1200</td>
<td>0.0343</td>
<td>0.0402</td>
<td>0.1563</td>
<td>0.1551</td>
<td>0.1202</td>
<td>0.1455</td>
</tr>
<tr>
<td>Poltava</td>
<td>2010</td>
<td>0.0605</td>
<td>0.0434</td>
<td>0.0323</td>
<td>0.0311</td>
<td>0.1322</td>
<td>0.0284</td>
<td>0.1394</td>
<td>0.1241</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>0.0673</td>
<td>0.0987</td>
<td>0.0568</td>
<td>0.0738</td>
<td>0.1546</td>
<td>0.1291</td>
<td>0.1384</td>
<td>0.0613</td>
</tr>
</tbody>
</table>
5. Conclusion

Statistics of the refractive index of the troposphere for each of the seasons of the year can be described locally by Gaussian models. The values of the mean deviation and RMS are minimal for the winter and the maximum for the summer.

For the first time was shown the possibility of using to describe the time-dependent behaviour of the refractive index during the season based on nested semi-Markov processes in which the description in the phase state (the season) can be based on the use of finite Kravchenko functions for the summer and Gaussian functions for the other seasons.

A new approach to test hypotheses about the law of the random variable is proposed. It is based on the calculation of the deviation of the experimental values of the distribution function, converted by the inverse function which used for their approximation with respect to the linear. Evaluation of the significance of this deviation for different testable hypotheses was performed by Fisher's exact test.

Acknowledgements

This work was partially funded by a grant of joint works of RFPF (grant No. 12-02-90425) and the NAS of Ukraine, the financial support of the Task integrated NAS programs on scientific space research for 2012-2016 and contract No. DZ/467-2011 with SASII of Ukraine.

REFERENCES