Simulation of Vector Semibinary Homogeneous Random Fields and Modeling of Broken Clouds

S. M. Prigarin[1,2]* and A. L. Marshak[3]**

[1]Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch, Russian Academy of Sciences, pr. Akad. Lavrent'eva 6, Novosibirsk, 630090 Russia

[2]Novosibirsk State University, ul. Pirogova 2, Novosibirsk, 630090 Russia

[3]NASA, Goddard Space Flight Center (Code 613.2), MD 20771 Greenbelt, USA

Abstract—A vector-valued homogeneous random field is said to be semibinary if its single-point marginal distribution is a sum of a singular distribution and a continuous one. In this paper, we present methods of numerical simulation of semibinary fields on the basis of the correlation structure and the marginal distribution. As an example we construct a combined model of cloud top height and optical thickness using satellite observations.

INTRODUCTION

This paper is a continuation of [1−3] where numerical methods to simulate binary random fields on a plane were developed. These methods were used to simulate stochastic geometry of broken cloud fields. A complex approach is needed to construct more adequate models for a detailed study of radiation balance and radiation transfer in the atmosphere. Such an approach would make it possible to take into account statistical dependence between the various components describing the geometrical and optical characteristics of clouds. Thus, it is necessary to create numerical models of random vector fields of special type. In this paper, a relevant class of random fields is described and methods of simulation are proposed for reproducing correlation properties and marginal distributions. As an example we consider the numerical model of an optical-thickness-and-cloud-top-height system of a cloud field constructed on the basis of satellite measurements.

1. DEFINITIONS AND PROBLEM STATEMENT

Definition. A homogeneous (vector-valued) random field $\xi(x)$, $x \in R^d$, is said to be *semibinary* if its single-point marginal distribution F_ξ (i.e., the distribution of random quantities $\xi(x)$) is the sum of a continuous distribution F_1 in R^k and a singular distribution F_0 concentrated on some element ξ_0 (which we assume to be numerical for simplicity; in principle, however, it can be of any nature), and we write $F_\xi = F_1 + F_0$. The element ξ_0 is called a *singular value* of the semibinary field ξ. A random field defined by setting

$$1_\xi(x) = \begin{cases} 0 & \text{if } \xi(x) = \xi_0, \\ 1 & \text{otherwise} \end{cases} \tag{1.1}$$

*E-mail: smp@osmf.sscc.ru

**E-mail: Alexander.Marshak@nasa.gov

is called an *indicator field* (or simply *indicator*) of the semibinary field ξ with the singular element ξ_0.

In what follows, a marginal distribution always means a single-point marginal distribution. The marginal distribution of a homogeneous binary random process or field can be represented as a sum of two singular distributions (for instance, concentrated at points 0 and 1). For semibinary random functions, we replace one of the singular distributions by a continuous one describing, for instance, a signal, whereas the remaining singular distribution corresponds to there being no signal. Such random functions are often encountered in various applications (for instance, time series and precipitation fields, cloud fields, etc.). For a vector semibinary field, either there is no signal for all components, which corresponds to a singular value, or its value is distributed continuously in R^k. Thus, for instance, if there is a cloud above the surface at a point $x \in R^2$, then its characteristics can be specified by some vector $\xi(x)$ (describing, for example, the optical density or heights of lower and upper boundaries of the cloud). If there is no cloud, then $\xi(x) = \xi_0$.

In this paper, methods to simulate numerically vector semibinary homogeneous random fields are developed and simulation models of broken clouds are constructed based on these. First, we present a numerical model of a scalar semibinary field. Next, we describe a simple generalization to the vector case. To make the problem more rigorous mathematically, we formulate some requirements which a numerical random-field model must satisfy. We will require that the numerical model reproduces marginal distributions of a semibinary field and a spatial correlation function either of the field itself or of its indicator. The method of inverse distribution functions and its modifications seem most suitable for solving the problem formulated above. In this case, a numerical model of a semibinary field is constructed by a nonlinear transformation of a specially selected Gaussian random function, i.e., as a quasi-Gaussian field.

2. SIMULATION OF SCALAR AND VECTOR SEMIBINARY RANDOM FIELDS

2.1. The main idea behind an inverse distribution function method and its modifications tailored to simulate non-Gaussian random fields is the following (for fundamentals, see [4], and for details, [5]). First, a Gaussian random field $u(x)$ with zero mean, unit variance, and a correlation function $\rho(y) = \mathbf{E}u(x+y)u(x)$ is simulated. Next, a random process ξ is simulated via the formula

$$\xi(x) = f(u(x)). \tag{2.1}$$

This method is also called a method of (noninertial) nonlinear transformations of Gaussian functions. Such a method makes it possible to simulate non-Gaussian homogeneous random fields with any marginal distributions and to approximately reproduce the required correlation function. Random fields ξ satisfying relation (2.1) will be referred to as *quasi-Gaussian*.

The marginal distribution is provided by an appropriate nonlinear transformation f. For the classical inverse distribution function method, a transformation f is monotone nondecreasing, that is,

$$f(u) = F^{-1}\Phi(u), \tag{2.2}$$

where Φ is a standard normal distribution function, and

$$F^{-1}(\alpha) = \inf\{z : F(z) > \alpha\},$$

which coincides with the inverse function in the case of a continuous distribution. A problem arising here is that an infinite number of transformations f can be constructed to provide the required marginal distribution, for instance,

$$f(u) = F^{-1}\big(g(\Phi(u))\big),$$

where g is an arbitrary mapping of the interval $[0, 1]$ into itself preserving the Lebesgue measure. In particular, a monotone nonincreasing transformation has the form

$$f(u) = F^{-1}\big(1 - \Phi(u)\big).$$

A covariance function of the random field in (2.1) is represented as

$$\mathbf{E}\xi(x)\xi(y) = r(x,y) = R_f\big(\rho(x,y)\big), \tag{2.3}$$

$$R_f(\rho) = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} f(u)f(v)\varphi_\rho(u,v)dudv,$$

where

$$\varphi_\rho(u,v) = \left[2\pi\sqrt{1-\rho^2}\exp\left(\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}\right)\right]^{-1}$$

is the density of a 2D Gaussian vector with zero mathematical expectation, a unit variance of the components, and a correlation coefficient between the components equal to ρ. Thus, for the random field in (2.1) to have a given covariance function of the form

$$\mathbf{E}\xi(x)\xi(y) = r(x,y),$$

the Gaussian field $u(x)$ must have the correlation function

$$\rho(x,y) = R_f^{-1}\big(r(x,y)\big). \tag{2.4}$$

Here we also have serious problems associated with the numerical inversion of (2.4) and possible lack of positive definiteness in the calculated function ρ (these problems are discussed in detail in [5]). Therefore, when solving many applied problems, we can only approximately reproduce the covariance structure of a simulated non-Gaussian field.

Thus, when using directly the method of nonlinear transformations of Gaussian functions to simulate semibinary homogeneous random fields, it is not clear how to find a nonlinear transformation such as in (2.1) and it is difficult to calculate correlations of the form (2.4). To simplify the situation, we propose that, instead of solving the previous simulation problem for a semibinary homogeneous field with a given 1D distribution and a given correlation function, we solve a simulation problem for a 1D semibinary field with a given 1D distribution and a given correlation function of the indicator field in (1.1). This approach seems to be reasonable for solving various applied problems because it simplifies the calculation algorithm, requires less preliminary information about the object being simulated, and makes it easier to adjust the model.

As a simulation algorithm, consider the following procedure. At the first stage, we simulate an indicator (binary) field $1_\xi(x)$ by using a Gaussian function $u(x)$, setting

$$1_\xi(x) = \begin{cases} 0 & \text{if } u(x) \in A_0, \\ 1 & \text{if } u(x) \in A_1, \end{cases} \tag{2.5}$$

where

$$A_0 \cup A_1 = (-\infty, \infty), \qquad A_0 \cap A_1 = \emptyset.$$

In essence, this is in fact the same method of nonlinear transformations. The preliminary simulation of the indicator field makes it possible to control more effectively the quality of simulation of the structure of a set on which the field takes values from a continuous distribution F_1.

For indicator cloud fields, in [1] it was proposed to use the following two models:

$$A_0 = (-\infty, d), \quad A_1 = [d, +\infty), \quad d \in (-\infty, +\infty), \tag{2.6}$$

and

$$A_0 = (-d, d), \quad A_1 = (-\infty, d] \cup [d, +\infty), \quad d \in (0, +\infty). \tag{2.7}$$

These were used in [2] to simulate random binary textures. In [3], it was shown that the models of stochastic structure of broken clouds based on (2.5)–(2.7) are in a good agreement with the results of ground-based and satellite observations. The parameter d is determined by averaging the indicator field $m_\xi = \mathbf{E}1_\xi(x)$ using the relations

$$m_\xi = 1 - \Phi(d)$$

for model (2.6), and

$$m_\xi = 2(1 - \Phi(d)) = 2\Phi(-d)$$

for model (2.7). The calculation of transformations of correlation functions (2.3), (2.4) is also significantly simplified (see [2] for more details). For (2.6), we have

$$R(\rho) = \Phi(-d) - 2T(d, a), \tag{2.8}$$

and for (2.7),

$$R(\rho) = 4\Phi(-d) - 4\{T(d, a) + T(d, 1/a)\}. \tag{2.9}$$

Here T is the Owen function,

$$T(d, a) = \frac{1}{2\pi} \int_0^a e^{-d^2(1+u^2)/2} \frac{du}{1+u^2}, \qquad a = \sqrt{\frac{1-\rho}{1+\rho}}.$$

After an appropriate model of an indicator field satisfying (2.5) has been chosen, at the second stage we simulate a realization of a semibinary field by setting

$$\xi(x) = \begin{cases} \xi_0 & \text{for } 1_\xi(x) = 0, \\ F_1^{-1}\Phi_1(u(x)) & \text{for } 1_\xi(x) = 1, \end{cases} \tag{2.10}$$

where $u(x)$ is a numerical realization of the Gaussian function with which the indicator field $1_\xi(x)$ was constructed, and Φ_1 is the distribution function of a standard normal random quantity under the condition that its value belongs to the set A_1. Expression (2.10) corresponds to the inverse distribution function method and provides the required 1D distribution for the semibinary field $\xi(x)$. For models (2.6) and (2.7), we have

$$\Phi_1(z) = \int_d^z \phi(z)dz/C, \qquad C = \int_d^\infty \phi(z)dz,$$

where ϕ is the density of a standard normal distribution.

2.2. Assume now that a homogeneous semibinary random field $\xi(x)$ is vector-valued, that is,

$$\xi(x) = [\xi_1(x), \ldots, \xi_n(x)],$$

and that the marginal distribution $F_\xi = F_0 + F_1$ of a field ξ, that is, an n-dimensional distribution of the vector $\xi(x)$, is known. In this case the above numerical simulation algorithm can be modified in the following way. At the first stage, again we use a Gaussian function $u(x)$ to simulate an indicator field $1_\xi(x)$. At the second stage, we construct a realization of a vector semibinary field by setting

$$\xi(x) = \begin{cases} \xi_0 & \text{for } 1_\xi(x) = 0, \\ \eta(u(x), x) & \text{for } 1_\xi(x) = 1. \end{cases} \tag{2.11}$$

Here $\eta(u(x), x)$ are n-dimensional random vectors simulated independently for various values of the second argument, with distributions $F_\eta(z_1, \ldots, z_n|u)$ defined by u. The distributions $F_\eta(z_1, \ldots, z_n|u)$ are chosen so that

$$\int F_\eta(z_1, \ldots, z_n|u)\phi(u)\, du = F_1(z_1, \ldots, z_n). \tag{2.12}$$

Formulas (2.11) and (2.12) provide the required marginal distribution of the simulated field, and the spatial structure of the field is determined by the correlation structure of the Gaussian function u. For the complex models of broken clouds considered below, the following method was used to construct

random vectors η. First, one component of the vector was simulated by the inverse distribution function method

$$\eta_k(u(x), x) = \left[F_1^{(k)}\right]^{-1} \Phi(u(x)), \tag{2.13}$$

Next, other components of the vector $\eta(u(x), x)$ were simulated by using the conditional distribution $F_\eta(\eta_m, m \neq k | \eta_k)$ for a known value of η_k. Here $F_1^{(k)}$ denotes the continuous component of the distribution of the kth component ξ_k of a vector semibinary field ξ.

Remark. A characteristic feature of the above simulation algorithm is that for the components of a semibinary field that differ from the component numbered k, the obtained realizations are much noisier than the real fields with which the numerical model is constructed (see Sec. 3 below). In fact, this is due to (2.11) and the fact that the random vectors $\eta(u(x), x)$ are independent for different x (at fixed values of $u(x)$). Randomization with respect to the index k, for which (2.13) is used, does not improve on realizations. A possible solution of this problem is to introduce an additional spatial dependence between the vectors η, which would apparently make the construction of the model much more difficult. We may, on the contrary, use a simplified approach—where formula (2.13) is employed for all components of the vector η. In this case the smoothness of realizations of the components of a vector semibinary field is determined by the smoothness of realizations of a Gaussian function. However, only distributions for the field components themselves, but not their joint distribution, will be reproduced correctly. For a simplified model where all vector field components are simulated with the help of monotone transformations (2.13) of the same Gaussian function, cross correlations between the components take largest possible values.

## 3. COMBINED SIMULATION OF OPTICAL THICKNESS AND CLOUD TOP HEIGHT OF BROKEN CLOUD

The above method was used to develop a combined model of optical thickness and cloud top height of spatial fields of broken clouds. Such models are of considerable interest in studying the processes of radiation transfer and radiation balance in a cloudy atmosphere (see, for instance, [3] and references therein). The initial realizations of random fields used to generate a simulation stochastic model and to adjust its parameters are shown in Fig. 1. A realization of the generated model is presented in Fig. 2. The simulation model was constructed under the assumption on being homogeneous and isotropic for the observed field.

Formally, the problem consists in the following. The observed field can be represented as a realization of a 2D vector semibinary random field $\zeta(x) = [\zeta_1(x), \zeta_2(x)]$, with components ζ_1 and ζ_2 describing the
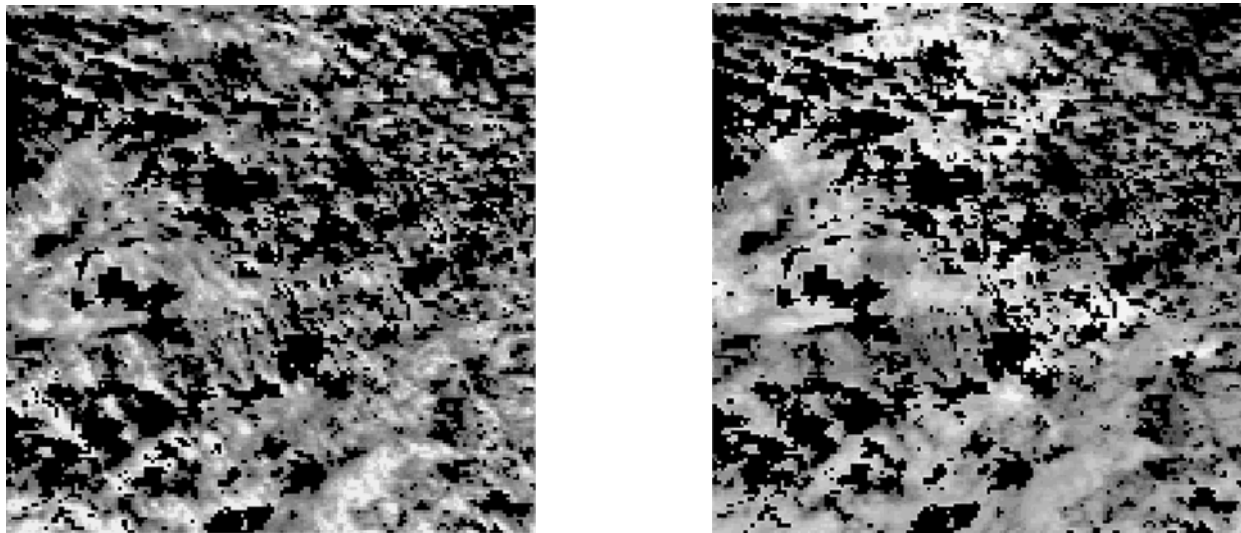
Fig. 1. Fields of optical thickness (left) and cloud top height (right) for clouds over a 160×160 km area at a resolution of 1 km. The images were retrieved from the measurements of MODIS spectroradiometer on board TERRA satellite.
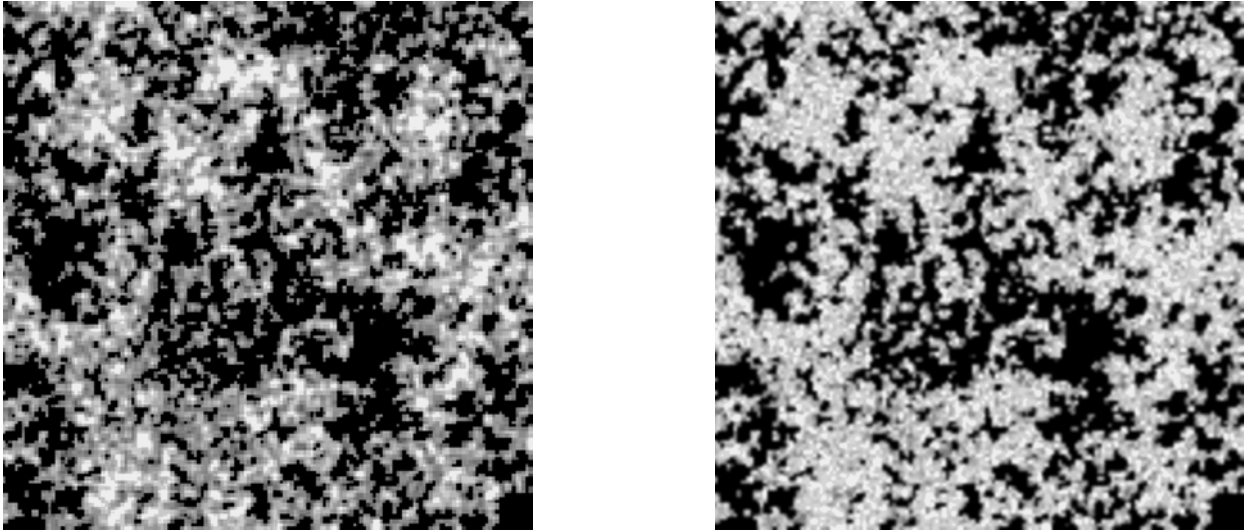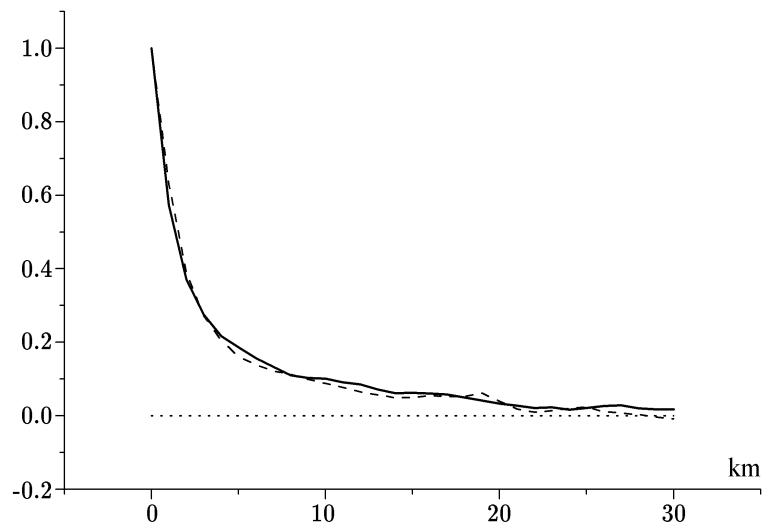
Fig. 2. Fields of optical thickness (left) and cloud top height (right) for clouds over a 160×160 km area at a resolution of 1 km. A realization of the stochastic cloud model.

optical thickness and cloud top height, respectively, and a 2D parameter x specifying coordinates in a plane. We need to construct a cloud model, that is, a model of a 2D vector semibinary random field $\xi(x) = [\xi_1(x), \xi_2(x)]$ whose characteristics and realizations would be similar to those of the observed field $\zeta(x)$.

Let us describe in more detail the procedure of generating the cloud simulation model using the general method presented in the previous section. At the first stage, we construct a numerical model of a Gaussian homogeneous isotropic random field $u(x)$ with zero mean, unit variance, and a correlation function $\rho(x)$. The correlation function is calculated by the formula $\rho(x) = R^{-1}(r(x))$ where $r(x)$ is the covariance function of an indicator field $1_\zeta(x)$, $r(x) = \mathbf{E}1_\zeta(0)1_\zeta(x)$ (it is estimated by the indicator of realization of the field $\zeta(x)$), and the function R is defined by expression (2.8) since we use model (2.5), (2.6). The parameter d is calculated from the equality $\mathbf{E}1_\zeta = 1 - \Phi(d)$ and the mean value $\mathbf{E}1_\zeta$, which in this case has the meaning of a cloud fraction, is also estimated by the available realization 1_ζ. A randomized spectral model proposed in [6] (more exactly, its modification for isotropic fields presented in [5, 7]) was used as a numerical model of a homogeneous Gaussian field $u(x)$. The numerical model of the homogeneous Gaussian field $u(x)$ constructed by using the indicator 1_ζ forms a basis for the simulation model $\xi(x) = [\xi_1(x), \xi_2(x)]$ of the optical-thickness-and-cloud-top-height system of a cloud. First, by using a realization of $u(x)$, the indicator field $1_\xi(x)$ was constructed from formulas (2.5) and (2.6); this is a simulation of the real indicator field $1_\zeta(x)$. Next, the components of a 2D semibinary field describing the cloud structure were simulated. An optical thickness field is taken as the first component $\xi_1(x)$. The component is simulated by the following nonlinear transformation of the Gaussian field (see (2.10)):

$$\xi_1(x) = \begin{cases} 0 & \text{for } 1_\xi(x) = 0, \\ F_{\xi_1}^{-1}\Phi_1(u(x)) & \text{for } 1_\xi(x) = 1. \end{cases}$$

Here $u(x)$ is a numerical realization of the Gaussian function used to generate the indicator field $1_\xi(x)$, Φ_1 is the distribution function of a standard normal random variable provided that its value belongs to the set $A_1 = [d, +\infty)$, and F_{ξ_1} is the continuous component of a distribution $\xi_1(x)$ estimated by the realization of $\zeta_1(x)$. Furthermore, using the available realization of $\zeta(x) = [\zeta_1(x), \zeta_2(x)]$, we estimate a conditional distribution of $\zeta_2(x)$ at fixed values of $\zeta_1(x)$. It is the conditional distribution that is used to simulate the second component $\xi_2(x)$ whenever the first component $\xi_1(x)$ is known. The generated stochastic cloud model is rather good in reproducing a joint empirical distribution of the components $[\zeta_1(x), \zeta_2(x)]$ of the observed field and a spatial correlation structure of the indicator field (see Figs. 3 and

Fig. 3. Normalized correlation function of the indicator field of the observed cloud shown in Fig. 1 (solid line) and of the generated one shown in Fig. 2 (broken line).
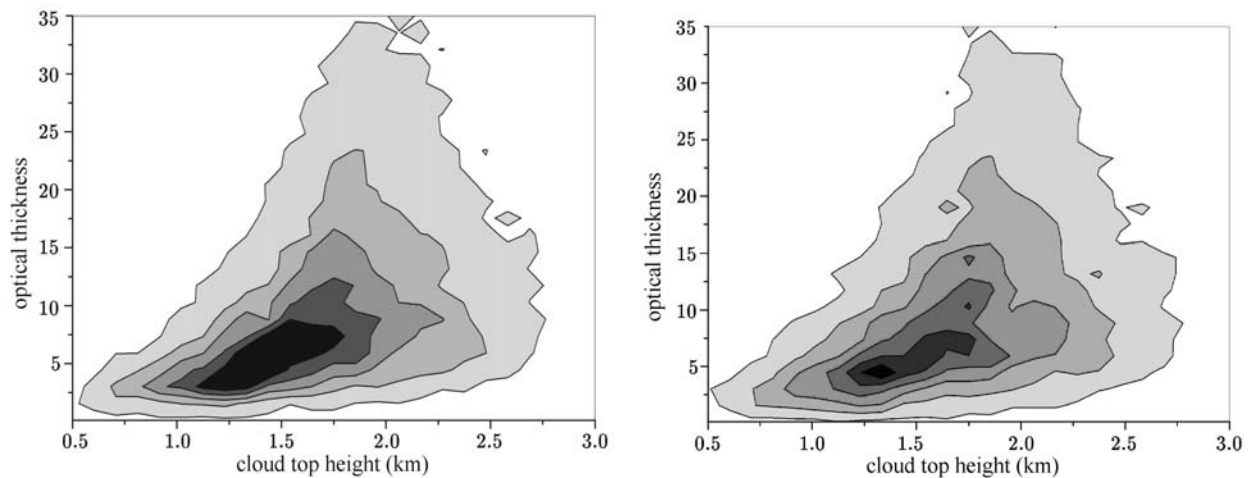

Fig. 4. Empirical densities of joint distribution of optical thickness and cloud top height obtained from satellite observations (left) and generated using the above cloud model (right); see Figs. 1 and 2, resp.

4). Another example of simulating the optical-thickness-and-cloud-top-height system is presented in Figs. 5 and 6.

It should be noted that, in principle, we can remove the assumption on being isotropic for the simulated random field, which was sort of strained for the cloud fields considered. In this particular case the assumption is used for a more precise estimation of distributions and covariances since the simulating algorithm in our example is constructed based on a single realization of a random field.

CONCLUSIONS

In this paper we have proposed a relatively simple and universal method for numerical simulation of a wide class of stochastic objects—homogeneous vector random fields whose single-point marginal distributions contain a singular component. The method is based on nonlinear transformations of Gaussian functions and intermediate simulation of the indicator random field. It is shown that the method with which we are coming up with is efficient for generating combined models of stochastic fields of optical thickness and cloud top height of broken clouds based on satellite observations. A software with the method and examples of simulation can be found at http://sergeim.prigarin.googlepages.com/Prigarin_CloudSim.pdf.
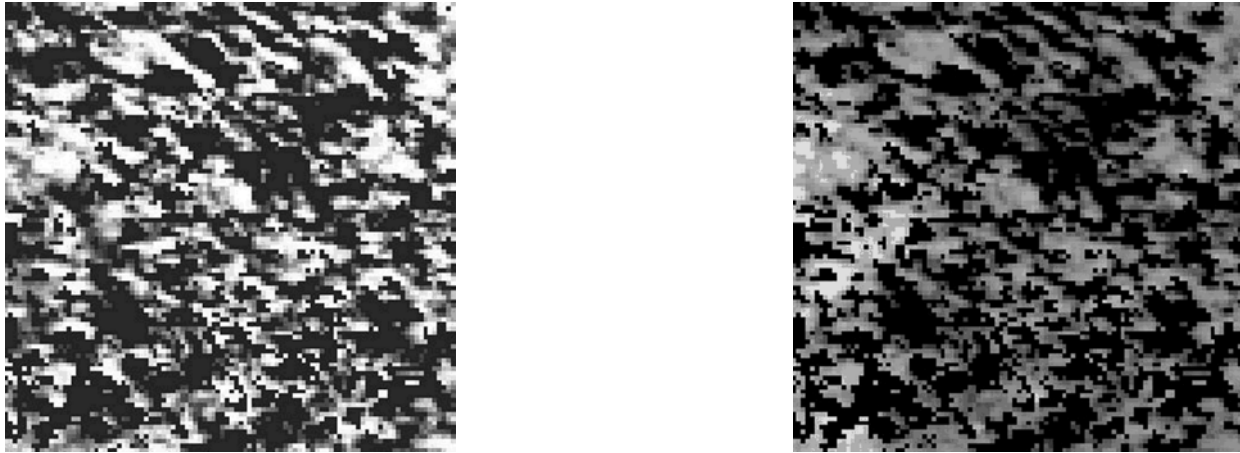
Fig. 5. Fields of optical thickness (left) and cloud top height (right) for clouds over a 100×100 km area at a 1 km resolution. The images were obtained from MODIS measurements.
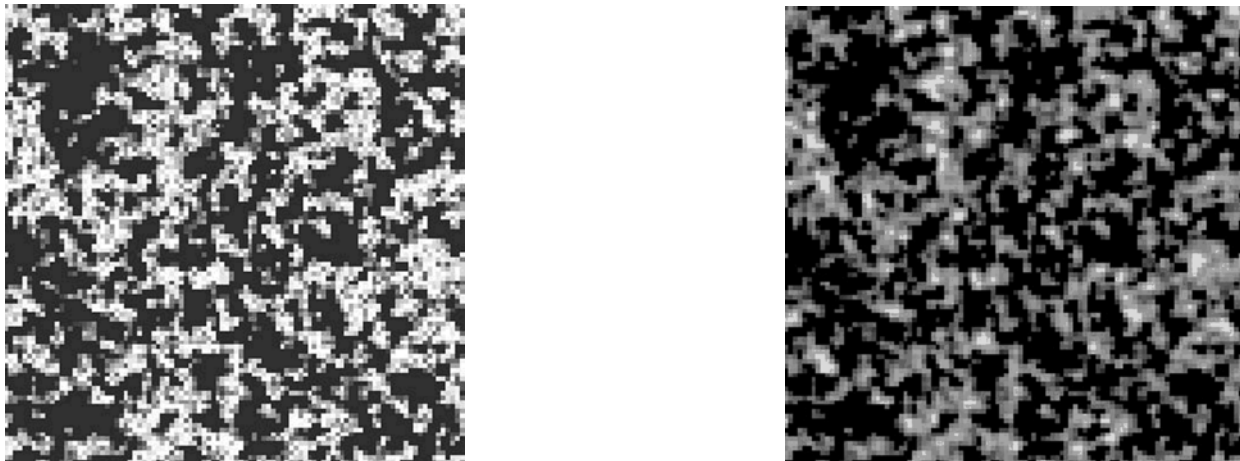
Fig. 6. Fields of optical thickness (left) and cloud top height (right) for clouds over a 100×100 km area at a 1 km resolution. Realizations of the stochastic cloud model.

ACKNOWLEDGMENTS

REFERENCES

1. Kargin, B.A. and Prigarin, S.M., Simulation of Cumulus Clouds to Investigate Processes of Solar Radiative Transfer in the Atmosphere by Monte Carlo Method, *Opt. Atmosf. Ok.*, 1994, vol. 7, no. 9, pp. 1275–1287.
2. Prigarin, S.M., Martin, A., and Winkler, G., Numerical Models of Binary Random Fields on the Basis of Thresholds of Gaussian Functions, *Sib. Zh. Vych. Mat.*, 2004, vol. 7, no. 2, pp. 165–175.
3. Prigarin, S.M. and Marshak, A.L., Numerical Simulation Model of Broken Clouds Adapted to Observation Results, *Opt. Atmosf. Ok.*, 2005, vol. 18, no. 3, pp. 256–263.
4. Piranashvili, Z.A., Some Questions of Statistical-Probabilistic Simulation of Random Processes, in *Voprosy issledovaniya operatsii*, Tbilisi: Metsniereba, 1966, pp. 53–91.
5. Prigarin, S.M., Methods of Numerical Simulation of Random Processes and Fields, Novosibirsk: Inst. Vych. Mat. Mat. Geofiz., 2005.
6. Mikhailov, G.A., Numerical Construction of a Random Field with a Given Spectral Density, *Dokl. Akad. Nauk SSSR*, 1978, vol. 238, no. 4, pp. 793–795.
7. Prigarin, S.M., *Spectral Models of Random Fields in Monte Carlo Methods*, Utrecht: VSP, 2001.