

Stochastic Hydrology: Is Scientific Understanding Growing?

*Mikhail Bolgov**

Institute for Water Problems, Russian Academy of Sciences, 3 Gubkina, 119333, Moscow, Russia

Abstract. Among many problems of stochastic hydrology, several major problems may be singled out. (1) The methodology problem – may fluctuation of hydro-meteorological values be considered within the framework of probabilities and random processes? Was this topic discussed after 1953? (2) One-dimensional probability distributions – is there progress? Are there new models? (3) Random Processes: Is Markovian property sufficient or more complex models with memory are needed? (4) Lack of stability resulting from climate changes: Is there progress in understanding the approaches to probabilistic forecasts?

1 Introduction

Probabilistic models and methods constitute the most important part of the land hydrology as a scientific discipline. Many hydrological processes and phenomena have been studied successfully over the last hundred years, after the Hazen and Sadler prominent works were published [21, 23] based on methods and models established within various sections of the probability and random processes theory. Interaction of land hydrology with technical disciplines contributed to the development of knowledge of properties and patterns of the hydrological process as well as formulation of these processes in the form needed to solve applied problems.

Within the applied hydrology, probabilistic models were generated making it possible to solve many applied problems, which, however, sometimes are rather empirical. Applied hydrology services to problems in the area of engineering systems' design and management meaning the need to take into account the requirements of these industries when formulating the hydrological results. These requirements are diverse, and yet, the ideas underlying the interaction of technical and natural sciences were based on the provisions of the reliability theory, defining the essence of engineering problems to be solved. The form of how problems solved in engineering directions were recorded was specific as well. We call it the regulations system. The certain rigidity of this system allows assuring the level of safety of the population, economy and territories acceptable to the community. There is also an understanding of difficulties arising when these methods are projected to new tasks associated, for instance, with taking into account environmental limitations and occurring climate changes. However, the attempts to replace the concept of reliability with the concept of acceptable risk were not too successful. Today, to further the Law on Technical

*Corresponding author: bolgovmv@mail.ru

Regulation [18], it is stipulated that the Technical regulations on the safety of buildings and facilities [19] based on the notions of the theory of reliability shall be in effect.

The stochastic direction in classical hydrology has also undergone the change of various views at models and methods used which were mainly borrowed from the theory of probability. A significant progress occurred during the post-war period as a result of introduction of the so called correlation theory of random processes in various forms. They couldn't do it without exaggerations, mostly ignoring properties of assessment of stochastic models.

The main results were obtained in the area of identification of optimal structure of models of random processes, taking into account, among other things, the asymmetric nature of distribution of run-off characteristics.

To summarize the Introduction, a conclusion can be made that the main methodology problem, i.e. whether the fluctuations in hydrological values within the theory of probabilities and random processes, and which models are preferable, are not discussed today. The discussion of this subject that took place in 1952 and 1953 has been buried in oblivion despite the fact that a significant amount of data has been accumulated, methods of the theory of probability have been developed, the arising water management problems have obviously become more complex and new environmental problems involving water management are waiting to be solved.

Out of a large number of stochastic hydrology's problems let's see only some, associated with current scientific problems and real-world applications.

2 One-dimensional probability distributions – Is there progress? Are there new models?

In stochastic hydrology, the problem of choosing the type of one-dimensional distribution and assessing its parameters has been the main issue starting from the first cases of application of probabilistic methods. Starting from the 40's of the last century, the notions of potential types of such distributions were formed that are meaningful through these days. The requirement of positivity and in general not being limited from above, as well as low complexity in accordance with the parsimony principle declared by V. Klemes. The principal approach to the construction of single-dimensional distributions was based on the transformation of one of the known laws (mostly, curves from the Pearson's family) or by means of some functional transformations. As an example, we can use the problem of determining the function of distribution of filling the reservoir that amounts to the solution of an integral equation of flow control [5]. Starting from the Kritski and Menkel [9] work, the three-parameter gamma-distribution used extensively in the domestic hydrology is well-known. Log-Pearson distribution, whose properties are close, is widespread in the U.S. hydrology.

Among totally new approaches, it is worth to mention the distribution functions that came to hydrology from the extreme value theory. Starting with B.V. Gnedenko works, the so called domains of attraction of distribution of extrema became known, to which the Weibull, Frechet and Gumbel [22] models correspond. They are used extensively in foreign practice of distribution of probabilities, so far are not used sufficiently in domestic hydrology.

Because the distribution of Kritski and Menkel was declared in the regulations to be primary [16, 17], the efforts of many science teams were directed at the development of methods to assess parameters, study selective properties of estimates [2, 14]. Out of last findings, we can mention the use of the method of linear moments to estimate parameters of these distribution, development of the truncation procedures, construction of two-dimensional distributions with marginal three-parameter densities, and alike.

In a number of cases it was impossible to find one distribution that would equally well describe the input data for the entire range. This is particularly visible when researching distributions in the area of extreme characteristics of run-off. One of the solutions consists in using mixed distributions. In [4] the authors discuss the mixture of two distributions: normal and power-series (Pareto's distributions). As a first component of the mixture the positive part of the normal distribution was used, while Pareto's distribution as a heavy tail served as the second component.

Data analysis on the maximum runoff in Primorye has shown that they were not able to find any particularly heavy tails. The review of the use of distributions with heavy tail is presented in the monograph [12].

Because the low and high runoff formation is affected by different factors, it would be feasible to exclude low runoffs from this discussion, i.e. use the so called truncated distributions. Out of relatively new results, it's worth mentioning the three-parameter gamma-distribution at an arbitrary point.

3 Random Processes: Is Markovian property sufficient, or more complex models with memory are needed?

In hydrology, an important place belongs to the problem of determining the type of the stochastic model of the process under investigation and the estimation of its parameters.

Research performed in the last decades has shown that in many cases it's acceptable to study the runoff as Markov's process. To describe fluctuations of river runoff with such approach the following methods will be used:

- Solution of Markov's equation as bilinear expansion in orthogonal functions.
- Stochastic differential equations in the form of Ito or Stratonovich.
- Diffusion equations by Fokker-Planck-Kolmogorov (FPK).
- Replacement of a variable in one of linear models to switch to values with significantly non-Gaussian one-dimensional distributions.

Markov process, by definition, is defined by the two-dimensional law of distribution, and therefore the main task of stochastic river runoff modelling within Markov approach is to take into account the non-Gaussian marginal distributions. The problem can be solved either by means of replacing a variable in the two-dimensional normal law of distribution, or, as was said earlier, by solving the Markov equation in the form of polynomials orthogonal expansion [15], or by means of calculation coefficients in stochastic differential equations.

For a number of years, the following problem was discussed in hydrologic research papers: which method gives us the two-dimensional density that best meets the empirical data. For no-lake rivers D.Ya. Ratkovich shown that the best model was the one obtained by I.O. Sarmanov by replacing the variable in two-dimensional density of evenly distributed random values, represented in the form of bilinear Legendre [10, 13] polynomial expansion.

It was relatively recently that the solution of the Markov equation was obtained in the form of bilinear polynomials orthogonal expansion for the three-parameter gamma-distribution giving symmetric two-dimensional density with a linear regression equation. Calculation of the required orthonormal polynomials is done by the method of Gram-Schmidt or thogonalization using their explicit representation through the moments of weighting function. In this case, the three-parameter density of distributions of Kritski and Menkel serves as the weighting function.

Fig. presents an example of regression equations for the two-dimensional law of distribution obtained using various methods with the same one-dimensional distributions of Kritski and Menkel.

Some of the recent years results in the area of description of patterns of long-term flow fluctuations were obtained using the methods utilizing the solution of stochastic differential equations. In the work of V.V. Kovalenko, V.I. Naidenov and a number of other authors [7, 8, 11], inflow fluctuation characteristics are obtained by way of solution of Fokker-Planck-Kolmogorov (FPK) diffusion equations. The models involved allow obtaining the probabilistic solutions fairly easily with various boundary conditions. The transition probability density $f(s,x; t,y)$ satisfying the Markov equation also satisfies the so called backward Fokker-Planck-Kolmogorov equation that takes the following form (Kovalenko, Sarmanov, 1978) [6]:

$$\frac{\partial f}{\partial s} = -a(s,x) \frac{\partial f}{\partial x} - \frac{b(s,x)}{2} \frac{\partial^2 f}{\partial x^2} \quad , \quad (1)$$

where $a(s,x)$ is called a drift coefficient, and $b(s,x)$ is called a diffusion coefficient (diffusion code).

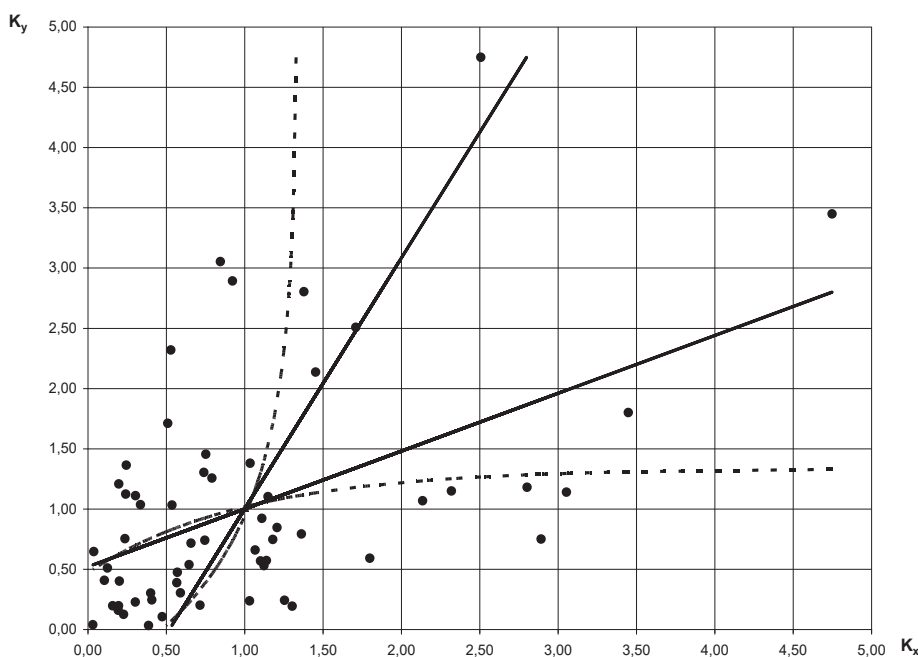


Fig. Regression lines for the joint distribution of adjacent values of annual inflow to the Lake Chany, calculated using a linear model 1 (full lines) and nonlinear model 2 (dotted lines); the dots are observed values of inflow.

The stationary solution of the FPK equation is described as:

$$p(x) = \frac{c}{b(x)} \exp \left[2 \int_0^x \frac{a(x)}{b(x)} dx \right] \quad , \quad (2)$$

For the stationary Markov process with a known two-parameter one-dimensional gamma-distribution that is most frequently used in stream runoff calculations the drift and diffusion coefficients in the FPK equation are calculated as [1, 3]

$$a(x) = -\mu(x-1), \quad b(x) = \frac{2\mu x}{\gamma} \quad (3)$$

Here, $\gamma=1/Cv^2$ is the parameter of one-dimensional distribution, and μ is the exponential index of correlation function.

As one of popular ideas in hydrological literature dedicated to long-term runoff fluctuations, the hypothesis of a self-similar structure of a time-series was discussed. It was shown empirically that in sequences of hydrometeorological characteristics, Hurst exponent estimated from selection [20] is much greater than 0.5. It means that the structure of the time-series is different from the white noise where the value $H=0.5$ is typical. However, for instance, analysis of rows of inflow to the Baikal lake and summer temperatures in its basin using the Hurst exponent identified an ambiguous situation. For the temperature row, where apparently a positive trend is present, the Hurst exponent $H=0.82$, while for the row of annual inflows to the Baikal Lake the Hurst parameter is 0.5, which, of course, may point at the simplest structure of the time-series, but in the structure of the inflow row, a group of low water years is identified, which is highly unlikely for the sequence of independent values.

Based on everything said, a conclusion can be made that the Markov model can be made refined by including more complex mechanisms of temporal variability, in particular to introduce various behavior of the process at different time periods. Technically this could be done if a Markov process is represented as Fokker-Planck-Kolmogorov equation allowing bistability. In V.I. Naidenov [11] work, an attempt was made to generalize the equation (1) for the case of several stable conditions, but the results were not developed any further. Multiple efforts to give hydrologo-geographical (hydrological and geographical) interpretation to the results of solution of the FPK equation were done in the works of V.V. Kovalenko and his students [7, 8].

4 Lack of stability resulting from climate changes: Is there progress in understanding the approaches to probabilistic forecasts?

Monitoring data have evidenced that in a number of regions of Russia, significant disturbances of homogeneity of runoff fluctuations, associated with climate changes, occur. One of new tasks of stochastic hydrology is probabilistic forecasting of runoff in the environment of climate changes.

Climate change scenarios adopted by the academic community do not cover all possible future conditions because our knowledge is limited. Understandings used in modelling (related to meteorology, ocean resources, hydrology, biology and alike) are imperfect. Scientific forecasting of complex, poorly studied phenomena that is of practical value, should be based on taking into account the diversity of existing contemporary knowledge and multiple-choice forecasts of global development. Obtaining such estimations is possible, for instance, using Bayesian approach taking into account chances of realization of a certain scenario. The main difficulty is to assess the chances of probabilities to implement a certain scenario. What is meant here is the attempts to distribute the chances among forecasts characterized by a great uncertainty. If the scenarios adopted by the academic community have no obvious advantages, then it would be safe to assume that their future chances to be implemented are equal, and the differences between them may be taken into account, for instance, by introduction of distribution of errors of verification and other modelling mistakes.

It should be noted that the existing hydrologic rows are too short to construct new probabilistic models by statistical methods only. In the second section of the article, stochastic models with several stable conditions were mentioned, but in case of characteristics of long-term runoff fluctuations the use of such models requires new physically-justified hypothesis. Contemporary hydrology has no such hypotheses of ar.

An acceptable hypothesis means that there are only two process states, and we can assume that with probabilities n_1/N and n_2/N the system may be in one of them. Here $n_1 + n_2 = N$, where N – the total duration of heterogeneous sample, consisting of two parts.

Regarding future runoff fluctuations, we can only predicate that with certain probabilities the system may be in one of the mentioned states. In this case we can assume that the expectation of the forecasted process is a mixture of two distributions, and the final stage of estimation (forecasting) consists in calculation of forecasted density using the total probability formula (Bayesian formula).

5 Conclusion

An answer to the question in the title of this article is definitely positive. In contemporary hydrology, new probabilistic models to improve the existing conclusions and results are developed, and approaches are offered allowing solving new tasks associated mainly with the change in conditions of how hydrological phenomena and processes are formed for different reasons.

Initial data limitation and changes occurring in natural and social environments, despite the significant progress of tasks of “global” modeling do not call off and probably won't call off in the future the need to develop probabilistic ideas and approaches. Only based on probabilistic approaches, it's possible to forecast runoff distribution taking into account the key sources of uncertainty in solving contemporary hydrology's problems.

This study was partially supported by the Russian Foundation for Basic Research, project no. 17-29-05108.

References

1. Blokhinov E.G., Sarmanov O.V. Gamma-correlation and its use in calculations of long-term river runoff control // Tr.GGI.-issue 141 L., Gidrometeoizdat. pp.52-75. (1968)
2. Blokhinov E.G. Distribution of probabilities of river runoff values, Moskva, Nauka, p. 169 (1974)
3. Bolgov M.V. Markov processes in the task of forecasting the levels of a landlocked body of water, *Meteorol. Giidrol.* **11** (2005)
4. Bolgov M.V., Pisarenko V.F. *Vodn. Resurs*, 26(6), pp.710-721. (1999)
5. Kartvelishvili N.A. The theory of probabilistic processes in hydrology and river runoff control, L. Gidrometizdat, p. 292 (1967)
6. Kovalenko I.N., Sarmanov O.V. A brief course of the theory of random processes. – Kiev, Vishchashkola., p. 262 (1978)
7. Kovalenko V.V. Hydrologic processes modelling, Gidrometeoizdat, S-Pb., p. 256 (1993)
8. Kovalenko V. V., Viktorova N. V., Gaidukova E. V. Hydrologicprocesses modelling – SPb.: izd. RGGMU, 559 p. (2006)
9. Kritskii S.N., Menkel' M.F. On the research techniques of random fluctuations of river runoff, TrudyGGI, Issue 29,Gidrometeoizdat., pp. 3–32 (1946)

10. Muzylev S.V., Prival'skii V.N., Ratkovich D.Ya. Stochastic models in applied hydrology - M.: Nauka, 174 p. (1982)
11. Naidenov V.I. Nonlinear dynamics of inland surface waters. Moskva, Nauka, 318 pp. (2004)
12. Pisarenko V.F., Rodkin M.V. Distribution with heavy tails: attachments to catastrophe analysis, Vychislitel'naya seismologiya. Issue No. 39, M., GEOS, pp 2-236 (2007)
13. Ratkovich D.Ya. Long-term fluctuations of river runoff - L.: Gidrometeoizdat, 255 pp. (1976)
14. Rozhdestvenskii A.V., Statistical methods in hydrology. Leningrad, Gidrometizdat, 424 pp (1974)
15. Sarmanov O.V. Investigation of stationary Markov processes by methods of expansion in eigenfunctions.//Tr. MIAN. pp. 239-261 (1961)
16. SNIP 2.01.14-83. Determination of design hydrological characteristics // Gosstroizdat SSSR. – M.: Stroiizdat, 36 p. (1985)
17. SP 33-101-2003. Determination of the basic design hydrological characteristics // Gosstroizdat Rossii. – M.: FGUP TSPP, 73 p. (2004)
18. Federal'nyi zakon "O tekhnicheskoi regulirovanii" dated 27.12.2002 No. 184-FZ
19. Federal'nyi zakon "Tekhnicheskii reglament o bezopasnosti zdani i sooruzhenii" dated 30.12.2009 No. 384-FZ
20. Hurst G. Neil. General Description of the River and the Use of Its Waters. M., Foreign Literature Publishers., 327 pp. (1954)
21. Hazen A. Storage to be Provided in Impounding Reservoirs for Municipal Water Supply. Proceedings of the American Society of Civil Engineers, 1913, Vol. 39, Issue 9, Pg. 1943-2044
22. Embrechts, P. Kappelberg, C. Mikosch, T. Modelling Extremal Events for Insurance and Finance, Springer (1997)
23. Sudler, C.E., Storage required for the regulation of streamflow. Trans. Amer. Soc. Civil Engrs., **91**, 622-660 (1927)