A Study on Multi-Component Flows in Gas Transmission Systems
(Flow and Measurement Processing Models)

N.A.Kislenko[1], M.G.Sukharev[1,2], A.S.Kazak[1], E.V.Fomina[1]

[1]Science Research Institute of Economics and Management Organization in the Gas Industry
(NIIgazekonomika LLC), Moscow, Russia
[2]Gubkin Russian State University of Oil and Gas (National Research University), Moscow, Russia

Abstract — **The focus of the study is on multicomponent natural gas flows through gas transmission systems (GTSs). The key objective is to determine the natural gas composition in each GTS branch. The initial data for calculation include measured natural gas components for all GTS metering points. The obtained measurements are considered to be random values due to instrument errors. Heat (calorific) value can be considered instead of natural gas components. A mathematical model has been developed for the calculation of natural gas composition for each pipeline. The gas composition changes at the joint points of the system. The model takes into account the irreversibility and non-equilibrium properties of mixing processes. The model is based on the well-known method of mathematical statistics, which is also known as the maximum likelihood method. It allows converting the problem to the quadratic programming problem with equality and inequality constraints. The equality constraints are the mass conservation equations for each fluid component, and inequality constraints are the relations demonstrating the incompleteness of components mixing at the joint points. A calculation example is presented to illustrate both the recommended calculation method and the approximate algorithm based on heuristic considerations. The developed approaches and methods are used to support the decision-making of diverse technological problems related to variation in the natural gas composition by gas transportation direction.**

I. Introduction.
The significance of multi-component gas mixture flow studies

Natural gas from different fields varies in composition. Normally, methane is a basic component. Its share in the gas of Russian fields lies in a range of 90 – 98%. Natural gas also contains other hydrocarbons, such as methane homologs (ethane, propane, butane, etc.), carbon dioxide, nitrogen, water vapor, helium, hydrogen sulfide, etc.

In addition to gas and gas-condensate fields, the gas sources for the Unified Gas Supply System (UGSS) of the Russian Federation are the oil fields (associated gas), gas processing plants (GPP), and underground gas storages (UGS). The composition of gas supplied to the UGSS can vary significantly depending on its source. Rich gases, which are supplied to the UGSS from GPPs, oil fields and UGSs constructed in depleted oil fields, usually contain a higher share of heavy hydrocarbons, and therefore have a higher heat (calorific) value. Gas from some GPPs contains up to 12-18% of ethane and 16-18% of nitrogen and its composition differs greatly from gas coming directly from gas fields. Now, the ethane share in the gas from the main UGSS gas sources is about 3%, and the nitrogen share is about 1.3%.

Hydrocarbons of methane homologous series are a more valuable feedstock for gas chemical facilities than methane. The economic viability of producing them from natural gas and using in gas chemistry does not cause any doubt. A more serious problem is where to locate the gas chemical facilities. There are two possible options either to place them closer to the gas fields or the industrial centers. Each option has its pros and cons. Under the first option, the methane content in gas transported by the UGSS pipelines

is about 100%. The second option suggests long-distance transmission of rich gas.

The gas fields of the Nadym-Pur-Taz region are the main source of the UGSS. A Russian port terminal on the Baltic Sea is considered to be a site for a gas-chemical facility. The transmission of gas along this route will require the collection of rich gas from the fields and modernization of the existing gas transportation corridors. The effective possible solution to the arising problems should be based on the scientifically proven models of multi-component gas mixture flows in large-scale main gas pipeline systems.

The range of applications of such models is not limited to the problems of UGSS prospective development. The calculation of the distribution of natural gas components concentration is also necessary for operating control of the gas transmission system. The mutual settlements, when both supplying gas to domestic consumers and exporting it, are performed not by volumetric indices but by energy ones, given heat (calorific) value of the supplied product. The procedure of gas pricing in Russia is regulated by federal legislation. The price is usually set according to the volumes of deliveries (in 1 000 cubic meters) adjusted for heat value. The differences in consumer requirements to gas calorific value make the supplier interested in delivering high-calorific gas to certain consumers but this makes it necessary to organize the UGSS flow control.

Another problem of GTS operational control is the need for the gas to be supplied to meet the requirements of dew point temperature. The dependence of dew point temperature value on water vapor mass concentration in natural gas is regulated by Russian [1, 3] and international [2, 4] standards. The dew point temperature values have to be controlled when exporting gas. The dew point temperature value is determined by calculating the mixing processes that occur in convergent and divergent gas flows, where the water vapor is considered to be one of the transported fluid components. This means that the calculation of the dew point temperature values is based on the models of multi-component gas flows in the pipelines.

It is worth noting that all the developed methods can be applied both to gases and to fluid flows. These methods can be used to solve the problems of oils blending which is rather important for maintenance services of the main oil pipelines. The oil-related issues are not considered in this paper, the gas transmission system terminology will be used.

II. Technological aspects

Gas samples are taken to track the calorific value and gas composition at gas measuring stations (GMS). Gas composition analysis is made periodically, using sample cylinders, or continuously, if the metering unit is specially equipped [3, 4].

The data on the distribution of calorific value and gas composition over the pipelines are of practical interest. To increase the information reliability one should take account of all the measurements carried out at the gas measuring stations. This approach will allow considering the measurements interdependency, however special methods and mathematical and computer models will be required for its implementation.

Depending on the specific character of technological problems, the models of both two-component and multi-component gas flows may be needed. The two-component model is required for example a) to determine the dew point, or b) in the case, if natural gas consists of two basic components (methane and ethane), and the share of other components is low and does not affect the result. Firstly, we will analyze the situation b) to simplify the understanding of the method and to facilitate its presentation.

The basic model will refer to the two-component fluid flow. Let us agree to call the components methane and ethane. The values of ethane concentration by gas transmission direction are the values to be determined. Gas calorific values can be considered instead of concentration values. The methods for determining calorific values for the flowing gas mixture are based on the same principles. The models of three- and more component- mixtures are based on the same principles as the models of two-component mixtures, only the calculation formulas become somewhat more cumbersome. The three-component model will be required for example to control the distribution of methane, ethane and helium concentration or to simultaneously calculate the distribution of calorific value and dew point temperature by transmission direction.

Further, in this paper, we will consider the pipeline system of an arbitrary configuration, which has several sources (i.e. fields, points of product inflow from adjacent systems, UGS under withdrawal conditions, etc.) and several sinks (points of product delivery to internal consumers and for export, UGS under injection conditions, etc.). The model is developed under the assumption that the total fluid rate for each pipeline of the system is known, i.e. the gas mixture flow distribution was calculated in advance and is the initial data to determine the flow rate of each component. At some points in the system — points, where the gas measuring stations are located — the fluid composition, is determined, i.e. the mass or molar the concentration of ethane in the transported product. Conversion of molar concentration to mass concentration does not cause difficulties. We use mass concentration for the sake of clarity.

The measurements may have errors, which means that the value of measured concentration is a random value. During gas transmission, the concentration values usually change due to the fluids mixing at the pipeline joint points. The concentration value does not change along each pipeline, consequently, it is usually related to the pipeline, and does not depend on the measuring point location. Conversion of molar concentration to mass concentration does not cause difficulties. We use mass concentration for the sake of clarity. The task is to estimate the concentration values for each pipeline over the entire set of measurements.

III. Non-equilibrium of mixing processes (technological aspect)

The success of solving this problem depends significantly on how correctly the calculation scheme is built. Calculation of large gas transportation systems often involves the replacement of several parallel gas pipelines with one equivalent arc of calculation scheme. This method is quite justified while searching for the flow distribution of gas mixtures in the system. However, if the fluids of different concentrations are mixed at any joint point of the calculation scheme, we cannot consider the concentration values for all pipelines diverging from that point to be equal. The concentrations would be equal if the mixing processes were equilibrium ones. In actuality, however, these processes are not equilibrium ones. The process could be an equilibrium one if the concentrations in the pipelines outgoing from the joint point were equal. However, the complete mixture of flows is not observed. Gas flows at rather high speeds. Concentrations in the outgoing pipelines depend on the local configuration of pipelines at the joint point. It is not possible to take into account such technological details in the aggregated models because the geometric dimensions of such an area are incommensurably smaller than the lengths of the pipelines. Therefore, to calculate the concentrations in all gas pipelines of the GTS, the calculation scheme should be disaggregated. It is necessary to ensure the possibility of reflecting the presence of different concentrations on the arcs outgoing from the joint point. Let us suppose that two pipelines or pipeline corridors meet at the joint point, and the multi-pipe corridor is the output. This corridor could be represented by one arc in the aggregated scheme. The aggregated model will not be suitable to calculate the concentration values in the case of different gas compositions in the incoming lines.

The processes of gas flow mixing are the main subject of this paper. The mixing occurs for two reasons – diffusion and turbulent mixing. The second reason is the main one for assessing changes in gas composition at the pipelines junction points. We study these particular changes and not a change in the distribution of components over the pipeline cross-section. In other words, one-dimensional problems on graphs are of practical interest. Observations of changes in operating parameters of the industrial gas pipelines do not give grounds to ascertain the generation of heat when the natural gas components mix. Even if the heat is produced, this does not affect the operation. Therefore, the mixing processes of multi-component gas flows are considered to be isothermal.

All gas mixing processes are irreversible, at least because it is necessary to spend some energy to separate the mixture into components. Mixing of the natural gas components during its transportation is a nonequilibrium process because complete mixing is usually not achieved, which should be considered as an experimentally established fact.

The irreversible nonequilibrium processes are studied by thermodynamics, physical chemistry, and kinetic gas theory. Irreversible and nonequilibrium processes are often used in mastered technologies for the production, transportation, and processing of natural gas [5, 6]. The application range of such processes is continuously expanding. Let us note some recent publications related to gas pipeline transportation. The problems of natural gas and hydrogen mixture transportation are considered in [7 – 8] and those of natural gas and nitrogen mixture transportation are discussed in [9].

The schemes of gas flow mixing are usually considered in theoretical and technical thermodynamics [10 – 16] with a focus on the study of pressure and temperature parameters of the mixing flows. Indeed, it is important to evaluate the additional efficiency loss due to irreversible heat transfer between mixing gases and due to their unused pressure difference.

The kinetic theory of gases and, in many aspects, physical chemistry use the models related not to volumes, but molecular or atomic structures. They use the apparatus of quantum chemistry, statistical mechanics, and analytical dynamics [17 – 19].

The industrial pipeline systems considered in this paper and the physical processes in them do not require the use of the methodological and mathematical apparatus traditionally used to study irreversible and non-equilibrium processes. It is highly likely that they have no immediate predecessors in the scientific and technical periodicals.

IV. Two-component gas. The mathematical statement of the problem

Structure of a pipeline system will be represented by oriented (by default, all arcs are assumed to be oriented in the direction of fluid flow) graph $G = (V,E)$, where value V is a set of nodes, and E is a set of arcs. The arc in the aggregated scheme corresponds to the flow direction. The number of nodes is denoted by m and the number of arcs - by n. The nodes are divided into three groups: V_{in} – sources, V_{out} – sinks and joint nodes V_{joint}. We assume that the sources and sinks are connected to the graph by a single arc – outgoing for sources and incoming for sinks. This paper uses the terminology from classical monographs on graph theory [20, 21]. The fluid flow $\boldsymbol{\xi} = \{\xi_{ij}\}, (i,j)\epsilon E$ is set on graph G. It consists of two mixing components $\xi_{ij} = \zeta_{ij} + \eta_{ij}$. Vector $\boldsymbol{\xi}$ is considered to be deterministic, although, the information used in its calculation is not reliable. The effect of this uncertainty on the calculation results is going to be a subject of future research. Here ξ_{ij} is fluid flow by arc(i,j). The values ξ_{ij} comply with material balance equations at all joint nodes x_k

$$\sum_{x_j \in \Gamma^{-1}(x_k)} \xi_{jk} - \sum_{x_i \in \Gamma(x_k)} \xi_{ki} = 0, \ x_k \in V_{joint}. \quad (1)$$

Hereinafter $\Gamma(x_k)$ is a set of nodes, which the arcs outgoing from x_k come into, $\Gamma^{-1}(x_k)$ is a set of nodes, which the arcs incoming to x_k go out (Fig. 1). Vector $\boldsymbol{\xi}$ is a set of all flow rates $\xi_{ij}, (i,j) \in E$. The sets of methane flow rate

$\zeta_{ij},(i,j) \in E$ and ethane flow rate $\eta_{ij},(i,j) \in E$ are also the vector values $\boldsymbol{\zeta}, \boldsymbol{\eta}$. They must comply with the balance equations (1) changing $\boldsymbol{\xi} \rightarrow \boldsymbol{\zeta}$ and $\boldsymbol{\xi} \rightarrow \boldsymbol{\eta}$, respectively. The flow $\boldsymbol{\xi}$ is initially known, consequently, all arcs $(i,j) \in E$ could be oriented by the flow direction, which is considered to be already done, thus, $\xi_{ij} \geq 0, (i,j) \in E$.

The task is to find the flow $\boldsymbol{\eta} = \{\eta_{ij}\}, (i,j) \in E$. We can consider the mass concentrations r_{ij} instead of the values η_{ij} using the relation $\eta_{ij} = r_{ij}\,\xi_{ij}$. The source of the information for solving the problem is a set of the gas composition measurements, i.e., measurements of concentrations r_{ij}. The measuring points are attached to the graph arcs; the notation E^* is introduced for the set of such arcs. To indicate the measured ethane concentration, we also use the symbol "*". The resulting measured value r_{ij}^* consists of true (but unknown) value r_{ij} and the measurement error $\delta r_{ij},(i,j) \in E^*$. In the theory of errors, measurement errors are considered to be normally distributed quantities.

$$\delta r_{ij} \in N\left(0, \sigma_{ij}^2\right). \tag{2}$$

Symbol $X \in N(a, \sigma^2)$ means that the random value X has the normal distribution with mean a and variance σ^2. The instrument errors (or the measurement method errors) are characterized by variance σ_{ij}^2.. The error in the determination of ethane flow rate $\delta \eta_{ij}$ also has a normal distribution $\delta \eta_{ij} \in N\left(0, \xi_{ij}^2 \sigma_{ij}^2\right)$. According to the problem statement, one should find the distribution of ethane flow rate that to the greatest extent is consistent with the whole set of concentration measurements. The probabilistic statement of the problem normally suggests the assessment rather than the determination of the unknown quantities r_{ij}. Mathematical statistics recommend the maximum likelihood estimation (MLE) for the point estimation of unknown parameters. According to the MLE, the estimate is the maximum likelihood function, which in our case is the probability of the totality of all measured values. The MLE leads to the problem of conditional minimization of a quadratic function.

$$\sum_{(i,j) \in E^*} \left((\eta_{ij}^* - \eta_{ij})/\xi_{ij}\sigma_{ij}\right)^2 \rightarrow min \tag{3}$$

if two conditions are satisfied:
the condition of ethane material balance

$$\sum_{x_j \in \Gamma(x_k)} \eta_{kj} - \sum_{x_j \in \Gamma^{-1}(x_k)} \eta_{ik} = 0, x_k \in V_{joint} \tag{4}$$

and the condition of irreversibility of the components mixing process, which will be discussed later.

Problem (3), (4) is a linearly constrained quadratic programming problem.

V. CONSIDERATION OF NON-EQUILIBRIUM MIXING PROCESS

The optimization problem (3), (4) does not fully reflect the two-component gas flow process from the physical point of view. The condition of the process non-equilibrium should be added to condition (4). The mixing process of gases with different compositions is irreversible and non-equilibrium. We will consider again a mixture of methane and ethane. Let us formulate the non-equilibrium conditions of the mixing process for this gas.

The non-equilibrium nature of the process manifests itself in the fact that at different concentrations of the flows entering the node, the concentrations in the outgoing lines may also be not equal to each other, i.e., gases do not mix completely: the mixing process at the joint nodes does not lead to an equilibrium state. In a sense, the mixing process is similar to the heat transfer process. In the case of heat transfer, the temperature in the communicating vessels levels out over time, tending to a weighted average value. In the non-equilibrium process of mixing multicomponent gases, the concentrations approach the average value.
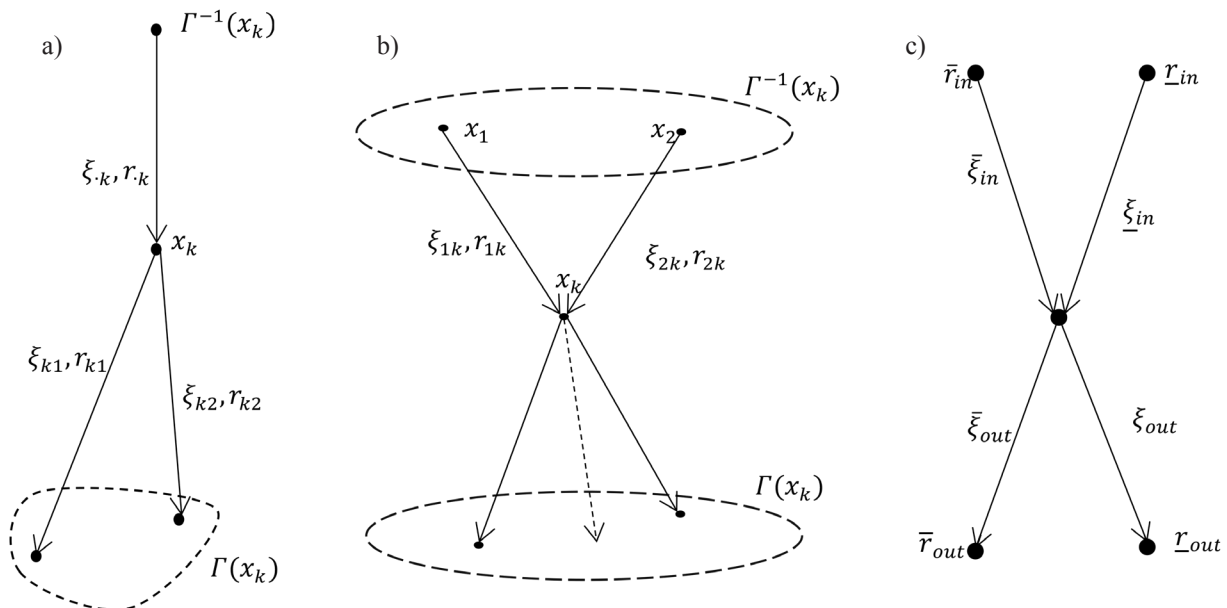

Fig.1. Schemes of diverging (a) and converging (b) flows at joint node x_k; notations for the non-equilibrium condition (c).

For simplicity, let us consider a scheme with two incoming and two outgoing lines. The parameters of the incoming lines will have the subscript *in*, and those of the outgoing lines – the subscript *out*, the parameters of the line where the concentration value of the 2nd component (ethane) is greater will be marked with an overscore and the parameters with lower concentration - with an underscore, i.e., $\max(r_{1k}, r_{2k}) = \bar{r}_{in}$, $\min(r_{1k}, r_{2k}) = \underline{r}_{in}$ (Fig. 1).

When the components mix completely, the concentrations in the outgoing lines (regardless of their quantity) are equal to $\bar{r}_{out} = \underline{r}_{out} = \rho$ independently of its quantity. The weighted average value ρ is expressed through the conditions at the inlet $\rho = \bar{\alpha} \cdot \bar{r}_{in} + (1 - \bar{\alpha})\underline{r}_{in}$, where $\bar{\alpha}$ is the share of the line where the ethane concentration is higher in the total (for two lines) gas flow rate. The following relations are necessary for the irreversible mixing process.

$$\underline{r}_{in} \leq \underline{r}_{out} \leq \bar{r}_{out} \leq \bar{r}_{in}. \tag{5}$$

These relations are necessary but not sufficient. They set the inequalities for extreme (maximum and minimum) concentration values at the outlet point but they do not take into account the relationship between the quantities of mixing components at the node inlet and outlet. Let the maximum concentration value in the outgoing lines be equal to the maximum concentration value in the incoming lines $\bar{r}_{out} = \bar{r}_{in}$. The flow rate in the outgoing line should not exceed the flow rate in the incoming line $\bar{\xi}_{out} < \bar{\xi}_{in}$ according to the physical non-equilibrium conditions. The similar physical relations should be satisfied for two, three and more outgoing lines.

We write these relations for the general case, denoting the number of incoming lines by N_{in}, and the number of outgoing lines by N_{out}. The consideration of lines with minimum and maximum concentration values will not be sufficient to make the analysis. We take any node and arrange the values of concentration in the incoming lines
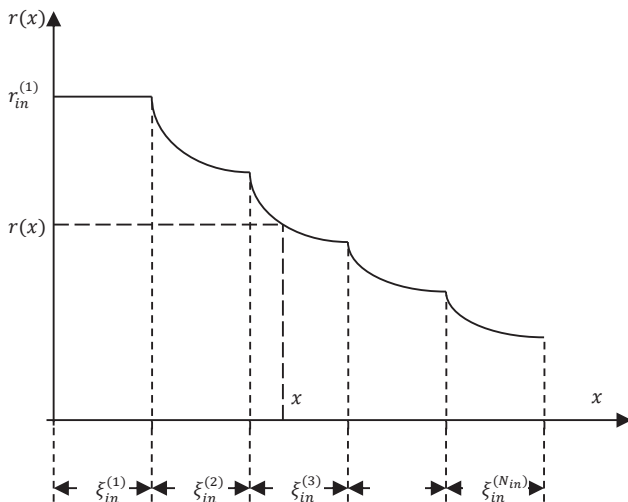

Fig. 2. Diagram of maximum possible concentrations (determined by the incoming lines of the joint node).

in decreasing order. A similar operation is performed for outgoing lines. Thus, two non-increasing sets of values are obtained

$$r_{in}^{(1)} \geq r_{in}^{(2)} \geq \cdots \geq r_{in}^{(N_{in})} \tag{6}$$

$$r_{out}^{(1)} \geq r_{out}^{(2)} \geq \cdots \geq r_{out}^{(N_{out})} \tag{7}$$

The superscript in brackets indicates the rank of the corresponding number in the sequence. It is obvious that $r_{in}^{(1)} = \bar{r}_{in}, r_{in}^{(N_{in})} = \underline{r}_{in}$ and $r_{out}^{(1)} = \bar{r}_{out}, r_{out}^{(N_{out})} = \underline{r}_{out}$. We use the ordered array of numbers (6) to construct the piecewise smooth curve of "limit concentrations" $r(x) = r_{in}(x)$ (Fig. 2). Thus, $r(x)$ is the maximum possible concentration of ethane in the gas mixture at the node inlet at the total fluid flow rate *x*.

$$r(x) = \begin{cases} r_{in}^{(1)}, if \ \ 0 \leq x \leq \xi_{in}^{(1)} \\ \frac{r_{in}^{(1)}\xi_{in}^{(1)}+r_{in}^{(2)}(x-\xi_{in}^{(1)})}{x}, \\ if \ \xi_{in}^{(1)} < x \leq \xi_{in}^{(1)} + \xi_{in}^{(2)} \\ \cdots \\ \frac{r_{in}^{(N_{in}-1)}\sum_{j=1}^{N_{in}-1}\xi_{in}^{(j)}+r_{in}^{(N_{in})}(x-\sum_{j=1}^{N_{in}-1}\xi_{in}^{(j)})}{x}, \\ \sum_{j=1}^{N_{in}-1}\xi_{in}^{(j)} < x \leq \sum_{j=1}^{N_{in}}\xi_{in}^{(j)} \end{cases} \tag{8}$$

Let us consider a point with coordinates $[x; r(x)]$, marked in Fig.2. It means that we will obtain the maximum concentration value with the inlet flow rate equal to x, if we add the total amount of gas supplied by the incoming line where the concentrations equal $r_{in}^{(1)}$ and $r_{in}^{(2)}$ and part of the gas amount supplied by the line where the concentration is $r_{in}^{(3)}$ at the flow rate $x - (\xi_{in}^{(1)} + \xi_{in}^{(2)})$.

In this case, the maximum mixture concentration value for flow rate x will be equal to

$$\frac{r_{in}^{(1)}\xi_{in}^{(1)} + r_{in}^{(2)}\xi_{in}^{(2)}x + r_{in}^{(3)}(x - \xi_{in}^{(1)} - \xi_{in}^{(2)})}{x}.$$

The distribution of flow rate and concentration values by outgoing lines will be acceptable if all points with coordinates

$$\left[\xi_{out}^{(1)}; r_{out}^{(1)}\right], \left[\xi_{out}^{(1)} + \xi_{out}^{(2)}; \frac{r_{out}^{(1)}\xi_{out}^{(1)}+r_{out}^{(2)}\xi_{out}^{(2)}}{\xi_{out}^{(1)}+\xi_{out}^{(2)}}\right], \ldots,$$

$$\left[\sum_{j=1}^{N_{out}}\xi_{out}^{(j)}; \frac{\sum_{j=1}^{N_{out}}r_{out}^{(j)}\xi_{out}^{(j)}}{\sum_{j=1}^{N_{out}}\xi_{out}^{(j)}}\right] \tag{9}$$

lie below the curve (8). The first of the points corresponds to the outgoing line with the maximum concentration, the second corresponds to two lines with concentrations $r_{out}^{(1)}$ and $r_{out}^{(2)}$, etc. Columns in Fig.3 represent the points. The last point with *x*-coordinate $\xi_{\Sigma} = \sum_{j=1}^{N_{out}}\xi_{out}^{(j)}$ and *y*-coordinate ρ always lie on curve $r(x)$. The distribution of concentrations (7) is admissible if the other points do not lie above $r(x)$. If we specify the *y*-coordinates of aggregate points (9) as $y^{(1)} = r_{out}^{(1)}, y^{(2)} = \frac{r_{out}^{(1)}\xi_{out}^{(1)} + r_{out}^{(2)}\xi_{out}^{(2)}}{\xi_{out}^{(1)} + \xi_{out}^{(2)}}, \ldots,$

$$y^{(N_{out})} = \frac{\sum_{j=1}^{N_{out}} r_{out}^{(j)} \xi_{out}^{(j)}}{\sum_{j=1}^{N_{out}} \xi_{out}^{(j)}}$$

then the admissibility condition will be written as follows:

$$y^{(1)} \le r\big(\xi_{out}^{(1)}\big), y^{(2)} \le r\big(\xi_{out}^{(1)} + \xi_{out}^{(2)}\big), \dots,$$
$$y^{(N_{out})} \le r\big(\sum_{j=1}^{N_{out}} \xi_{out}^{(j)}\big) \quad (10)$$

VI. PROBLEM-SOLVING METHOD

According to the problem statement, it is necessary to find the distribution of the component $\boldsymbol{\eta} = \{\eta_{ij}\}, (i,j)\epsilon E$, or, what is the same, the concentrations $\mathbf{r} = \{r_{ij}\}, (i,j) \in E$ that are most consistent with the whole set of concentration measurements. Using the maximum likelihood estimation (MLE), the only reasonable method of mathematical statistics for estimating unknown parameters, we obtained the problem of mathematical programming $(3 - 5)$. Let us denote by $n^*, m_{in}, m_{out}, m_{joint}$ the number of measuring points, sources, sinks, and joint points, respectively. We will start with an analysis of the simplified problem $(3 - 4)$, which is a quadratic programming problem with unknowns and m_{joint} linear constraints. There can be different cases depending on the values of $n^*, m_{in}, m_{out}, m_{joint}$. We will consider the simplest example – a system of three gas pipelines with one inlet and two outlets with 3 options of measured parameters. In option 1, concentrations are measured for each pipeline $n = n^* = 3, m_{joint} = 1$. Thus, we obtain the problem of quadratic function minimization of 3 variables at one linear constraint. From the constraint, one of the unknown values η_{ij} is expressed through the other two. The resulting problem on the absolute extremum of a quadratic function of two variables is reduced to the system of two linear equations with the same number of unknowns. While solving this problem, we obtain the estimates $\hat{\eta}_{ij}$ of unknown variables η_{ij}, which rely on all

Fig. 3. Verification of the fulfillment of the non-equilibrium conditions at the joint node.

three measurements. They are more justified than direct measurements η_{ij}^* of each of the values.

In option 2, ethane concentration is measured only on 2 pipelines (for example, r_{k1}^*, r_{k2}^* in Fig.1a). Now $n = 3$, $n^* = 2$, $m_{joint} = 1$, we have the problem of minimizing the quadratic function of two variables η_{k1}, η_{k2} with one linear constraint. The minimum of the likelihood function is found directly, and the obtained estimates equal the measurements $\hat{\eta}_{k1} = \eta_{k1}^*, \hat{\eta}_{k2} = \eta_{k2}^*$. The constraint is used to find the estimate of the missing unknown of the ethane flow rate along the arc entering node x_k. In option 3, we measure only one value, for example, η_{k1}^* (Fig. 1a). The maximum likelihood method enables us to obtain only a trivial result, i.e. the estimate of ethane flow rate by one arc $\hat{\eta}_{k1} = \eta_{k1}^*$.

The same kind of reasoning is carried out in the general case for any relationship between n, n^*, m_{joint}. It helps to reveal what results can be obtained with the existing system for measuring gas composition. The estimates of concentrations cannot be always obtained for all arcs of the graph. Everything depends on the number and location of measuring points. For a graph of an arbitrary structure, the determining value is $d = m_{joint} - (n - n^*)$. When $d > 0$, the desired estimates are refined because the mutual influence of the entire set of measurements is taken into account. If $d = 0$, the results of measurements directly serve as estimates of the concentration values. If $d < 0$, the constraint equations are not enough to estimate all non-measured concentration values r_{ij}. Graph G can have the subgraphs that meet the condition $d > 0$. The subgraph G_l (X_l, E_l) of graph $G(X, E)$ is the graph for which $X_l \subset X$ and for every node $x_k \in X_l, \Gamma_l(x_k) = \Gamma(x_k) \cap X_l$ [20, 21]. Consequently, quite a lot of measurements are made on the arcs of these subgraphs to refine the concentration estimates based on them. An algorithmic procedure for finding such subgraphs is proposed. Estimates are obtained as a linear function of measurements

$$\hat{r}_{ij} = \sum_{(k,l)\in E^*} \big[a_0^{ij} + a_{kl}^{ij} r_{kl}^*\big], \quad (i,j) \in E. \quad (11)$$

The quality of estimates is characterized by the variance. In the assumption (2) the variance of estimate $\hat{r}_{ij}$, following from the MLE, will be calculated by the formula

$$\mathbf{D}\hat{r}_{ij} = \sum_{(k,l)\in E^*} \big[\big(a_{kl}^{ij}\big)^2 \mathbf{D} r_{kl}^*\big], \quad (i,j) \in E \quad (12)$$

Let us note that there can be inconsistent data in the initial information due to its stochastic nature. Consider the following situation: at the junction of 3 pipelines (Fig. 1a), measurements are made on the incoming and one of the outgoing lines $r_{\cdot k}^*, r_{k1}^*$, with $r_{k1}^* > r_{\cdot k}^*$. Relationship (5), which in our case looks like $r_{k1} \le r_{\cdot k}$, should be satisfied to obtain the true concentration value. If the probability of inequality $r_{k1}^* > r_{\cdot k}^*$ is low, the result should be considered to be senseless. A possible reason could be the presence of
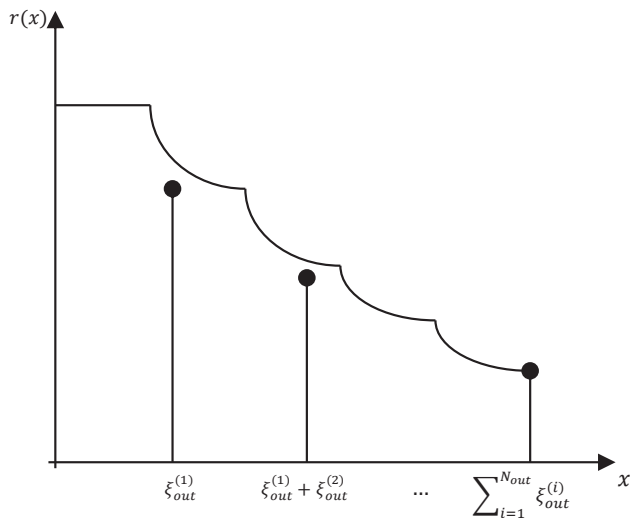
systematical measurement errors at one or both measuring points, which requires an audit of the devices or the qualifications of the personnel making the measurements.

Let us discuss the general case – the problem $(3 - 5)$. We paid so much attention to the simplified version (problem $(3, 4)$), because the procedure for numerically solving it is rather simple. Potentially, this solution can satisfy condition (5). Then the required result is achieved by small efforts, and further analysis is not required. Otherwise, a more complex mathematical programming problem is obtained, which, however, can be solved using standard optimization packages. While preparing the information to solve the problem, it is worthwhile to take into account the relations

$min(x, y) = (x + y - |x - y|)/2$,
$max(x, y) = (x + y + |x - y|)/2$,

and formalize search of $min(x_1, \dots , x_n)$ and $max(x_1, \dots , x_n)$ as algorithmic procedures.

VII. HEURISTIC ALGORITHMS OF LOCAL OPTIMIZATION

The procedures proposed above make it possible to obtain a solution (distribution of concentration values r over the arcs of graph G). Moreover, the set of arcs E of the graph splits into 3 noncrossing subsets $E = E' \cup E'' \cup E'''$. E' is a set of arcs with the estimates of concentration values for which $\hat{r}_{ij}, (i, j) \in E'$ takes into account the entire set of measurements affecting these estimates. The estimate of concentration values $\hat{r}_{ij}, (i, j) \in E''$ on the arcs E'' is equal to the concentration value measured on the corresponding arc $\hat{r}_{ij} = r_{ij}^*$, because due to limited information contained in the set of measurements, there are no other measurements that affect this estimate. E''' is a set of arcs for which the estimate of concentration $\hat{r}_{ij}, (i, j) \in E'''$ cannot be obtained under the existing system of measuring points in the considered GTS.

In practice, with insufficient measurements, it is worthwhile to have a picture of the concentration distribution, which is not necessarily rigorously substantiated but at least plausible. To this end, a heuristic algorithm is developed to sequentially look through the joint nodes, one at a time. Moreover, at each step, the calculation procedure is simple, and the amount of calculations is small, which is an attractive feature of the method.

Let us give some comments before the algorithm description. All nodes of the graph can be numbered so that the source node number for every arc is less than the sink node number. We order all arcs $(i, j) \in E$ and all nodes $x_k \in V$ of the graph by the numbers of the natural series. Assigning a number to an element – a node or an arc – we will say that this element is colored. We start numbering the nodes from the sources, assigning numbers $1, 2, \dots, m_{in}$ to them in random order. The joint nodes $x_k \in V_{joint}$ will be numbered as $m_{in} + 1, \dots, m_{in} + m_{joint}$, by following the rule – the next number is assigned to node $x_k \in V_{joint}$ after all nodes from set $\Gamma^{-1}(x_k)$ have been colored. After

numbering and coloring the node, we number and color all arcs outgoing from this node. The sinks are numbered as $m_{in} + m_{joint} + 1, m_{in} + m_{joint} + 2, \dots, m_{in}$ in any order. As a result, the numbers of nodes can increase only in the direction of flow. The source node number for each arc will be lower than the sink node number. The presented numbering scheme takes into account that graph G is an oriented graph, which does not have loops, i.e. oriented closed circuits. An example of numbering the graph nodes and arcs is presented in Fig. 4.

According to the numbering, the components of vector r are sequentially determined. With this in view, two heuristic algorithms of local optimization were developed.

Algorithm 1.

Step 0. Start of calculation. Color the nodes of set V_{in} because the information about gas composition in sources is included in the input data. Color the dangling arcs outgoing from the sources, assigning the appropriate concentrations to them.

Iteration: steps 1, 2, 3.

Step 1. Color the nodes, for which the incoming arcs are colored.

Step 2. Randomly select one of the colored nodes. Let this node be x_k. Determine concentrations r_{kj} at the arcs of the set $(k, j) \in \Gamma(x_k)$ outgoing from node x_k. If none of these arcs is included in the set E^*, then assign the same concentrations r_{kj} calculated from mass conservation condition (4) to arcs $(k, j) \in \Gamma(x_k)$, and color these arcs and node x_k. If one or several arcs $(k, j) \in \Gamma(x_k)$ (but not all arcs) belong to set E^* (i.e. the concentration values $r_{kj} = r_{kj}^*$ are measured), assign concentration values r_{kj}^* to these arcs and assign equal concentration values, calculated from condition (4)

$$r = \left(\xi_k r_k - \sum_{j \in E^*} r_{kj} \xi_{kj} \right) / \sum_{j \ E^*} \xi_{kj}$$

to the other arcs $(k, j), j \notin E^*$.

If all arcs from $\Gamma(x_k)$ are included in set E^*, i.e., $\Gamma(x_k) \in E^*$, all measured concentration values should be corrected by mass conservation condition (4) for ethane $r_{kj}^*, j \in \Gamma(x_k)$. Introduce the corrective multiplier $\lambda = \xi_k r_k / \sum_{j \in \Gamma^{-1}(x_k)} r_{kj}^* \xi_{kj}$, and assign the concentration values λr_{kj}^* to the arcs. Here ξ_k, r_k are the total fluid flow rate and ethane concentration at the inlet of node k.

All arcs $(k, j), j \in \Gamma(x_k)$ are colored. All nodes, for which all incoming arcs have already been colored, are also colored.

Step 3. Check if there are any uncolored arcs. If *yes*, go to step 1, if *no* –stop. The proposed algorithm takes into account that the values r_{ij} are small $r_{ij} \ll 1$, so it will be possible to limit the calculations by first-order terms in the calculations and assume, in particular, that fluid flow rates through the arcs cannot be changed when adjusting the concentration values.

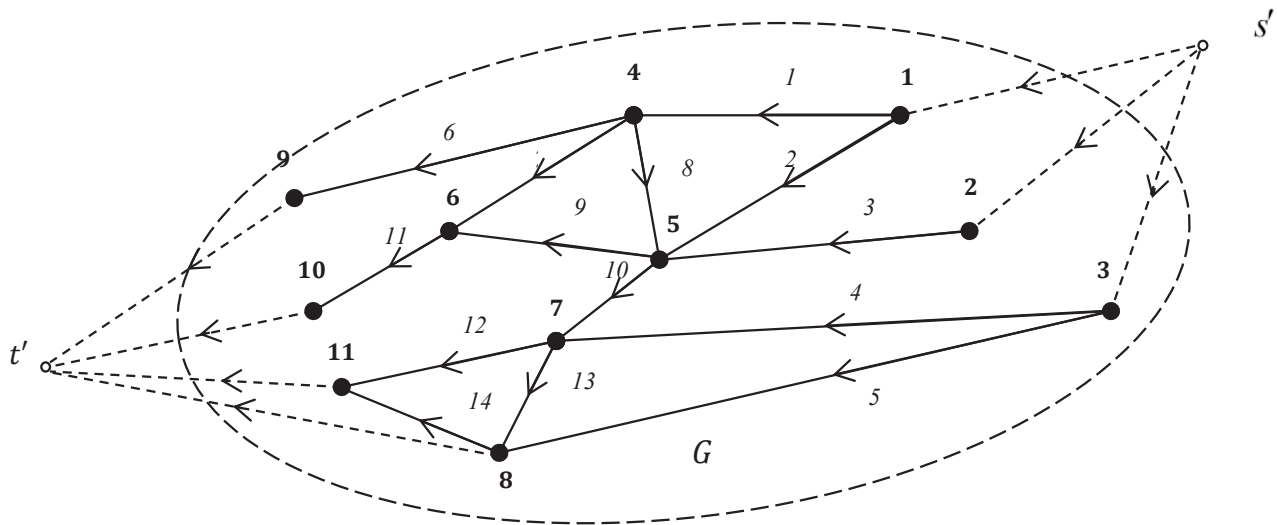*Note.* The order of algorithm realization can be reverse

Fig. 4. Graph of the system with fictitious source and sink.

if the elements are colored against the orientation of arcs. It is advisable if the data on concentration in sinks $r_{kj}, j \in V_{out}$ are included in the initial information.

The proposed algorithm gives an approximate solution that does not take into account the information on the entire set of measurements, both in sources and at other points of the system. The exact statement and solving the problem under standard assumptions about measurement errors were given above.

VIII. DECOMPOSITION METHODS

All ideas of algorithm 1 are generalized in algorithm 2 to be discussed later. We will give some definitions before describing the algorithm. We introduce the fictitious source s' by drawing arcs from it to each actual source and the fictitious sink t' by drawing arcs to it from each actual sink.

With the view to obtaining plausible and more precise numeric results than the results obtained with algorithm 1, we can also apply the decomposition methods. These methods allow reducing the search for solution r on graph G to a sequence of similar problems on its subgraphs. These problems are of smaller dimension and, consequently, are easily calculated. The decomposition methods are approximate and it is hardly possible to determine the extent to which they are accurate. Without going into detail, we will describe the idea of the methods.

The *minimal cut S* (hereinafter we use the term cut for simplicity) is a set of the graph arcs $S := (i,j) \in E$, without which the graph loses its connectivity. The property of connectivity loss will disappear after removing any arc from the set.

We can introduce partial ordering $\succ$ on the set of cuts assuming that $S_j \succ S_i$ if the numbers (the numbering is considered to be made as mentioned above.) of all arcs S_j or some of the numbers of arcs S_j are greater than numbers of all arcs S_i, and the other arcs from S_j belong to S_i (Fig.5).

We introduce the notation: the cut S divides G into

two connectivity components which will be denoted by $G_{source}(S)$, $G_{sink}(S)$, (the first includes the source and the second – the sink.), i.e. $G = G_{source}(S) \cup S \cup G_{sink}(S)$.

Adjustment of a solution by minimum cuts (Algorithm 2).

We sequentially proceed from cut S_{l-1} to cut $S_l \succ S_{l-1}$. . When doing so, we adjust the concentration values of arcs in set $(G_{source}(S_l) \cup S_l)/(G_{source}(S_{l-1}) \cup S_{l-1})$ using the ethane balance conditions as in algorithm 1, the transition from one node to another of a larger number was made. Then, the components of the solution on the subgraph (G_source $(G_{source}(S_l) \cup S_l)/(G_{source}(S_{l-1}) \cup S_{l-1})$ will not have the inverse effect on the solution on the subgraph $G_{source}(S_{l-1}) \cup S_{l-1}$.

IX. CALORIFIC VALUE DISTRIBUTION

It is easy to switch from the distribution of concentration values by an arc of the calculation graph to the distribution of heat (calorific) value. According to the State Standard 31369-2008 [22], the molar calorific value of gas mixture is calculated as a weighted average value by the gas
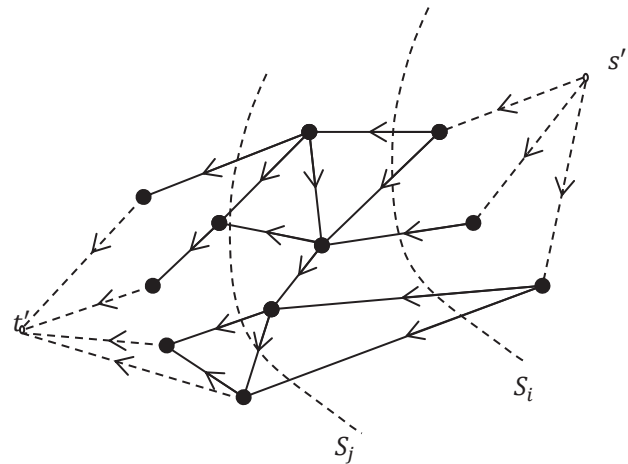

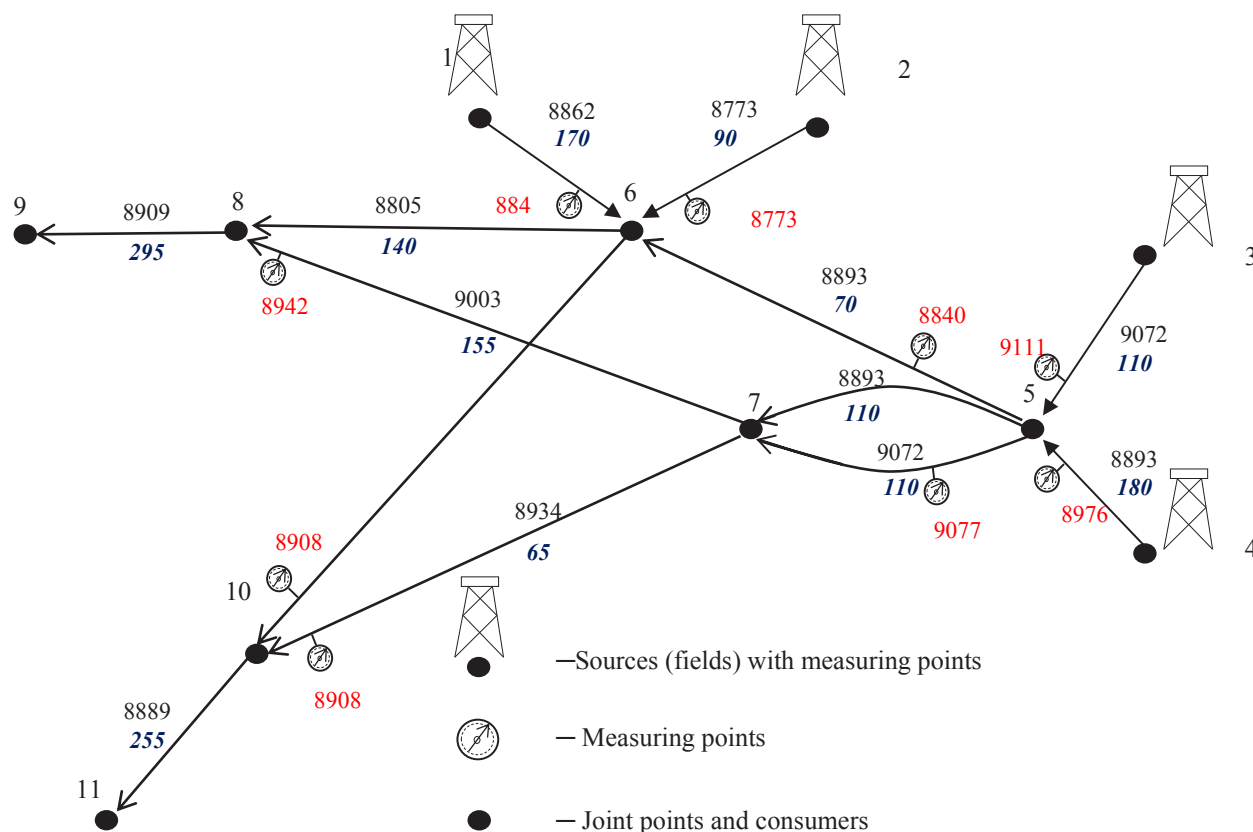*Fig. 5. The concept $S_j \succ S_i$.*

Fig. 6. Example. Distribution of gas volumetric calorific value (in kcal/cub m). 11 – number of node, 8936 – calorific value H_{ij} by arc, 65 – flow rates ξ_{ij} by arc (in MMcmd), 8910 – measured calorific values.

composition $\overline{H} = \sum_j x_j \overline{H}_j$, where x_j is the molar fraction of component j, and $\overline{H}_j$ is its calorific value. The values $\overline{H}_j$ for ethane and methane are 891,56 and 1562,14 $\frac{\text{kJ}}{\text{mol}}$, respectively. The mass calorific value $\widehat{H}$ is calculated by the molar one, using the formula $\widehat{H} = \overline{H}/M$, where M is the mixture molar mass. Volumetric heat value $\widetilde{H}$ is also proportional to value $\overline{H}_j$. Therefore, the calorific value can be calculated either by the concentration values or by the formulas mentioned above with changed notations. The calorific value plays a significant part in mutual settlements between gas suppliers and consumers, which is why this concept is used rather widely. A calculation example below clarifies the results in terms of calorific value.

X. CALCULATION EXAMPLE

By way of illustration, we will demonstrate the calculation of the calorific value distribution by transmission direction. Figure 6 presents the GTS structure, input data, and results of calculations.

Graph of the system includes n=13 arcs (transmission directions), $m = 11$ nodes (of which $m_{in} = 4$ sources, $m_{out} = 2$ sinks (consumers), $m_{joint} = 5$ joint nodes), *and* $n^* = 9$ measuring points. Therefore, $d = m_{joint} - (n - n^*) = 1$, and the number of measurements is sufficient to take into account their cross-impact and to make a system-wide estimation of unknown values.

The input data (flow rates, measured calorific values) and calculation results are demonstrated in the Table and some of them are given in Figure 6. $\widehat{H}_{ij}^{(1)}$ are the estimates of calorific value provided that condition (5) is met, $\widehat{H}_{ij}^{(2)}$ are the estimates obtained without meeting condition (5), $\widehat{H}_{ij}^{(3)}$ are the estimates obtained using the heuristic algorithm 1, $\hat{r}_{ij}$ are the estimates of concentration values in the main calculation with condition (5) satisfied. An analysis of the obtained results indicates that estimates $\widehat{H}_{ij}^{(2)}$ do not satisfy conditions (5) at node 5, while at the other nodes, conditions (5) are satisfied. Checking makes it possible to establish that in the case of solution $\widehat{H}_{ij}^{(1)}$ the sufficient condition (10) is also satisfied. In other words, simple models were used for the calculation, which did not include condition (10), but this condition appeared to be satisfied and further studies were not required. Therefore, one can hope that in many practical problems the solution to the optimization problem (3) subject to (4) and (5) will satisfy more stringent conditions (4) and (10).

The values of MLE criterion (3) are demonstrated in the last line of the Table. Value $\widehat{H}_{ij}^{(2)}$ is 14% less than the value $\widehat{H}_{ij}^{(1)}$. Such a large difference in the criterion values shows its sensitivity. The difference between vectors $\|\widehat{H}_{ij}^{(1)}\|$ and $\|\widehat{H}_{ij}^{(2)}\|$ is not large, whereas the difference between the values of the criterion for these vector arguments is considerable, i.e. if we do not allow for the irreversibility

Table 1. The initial data and calculation results

Arcs(i,j)	Flow rates ξ_{ij} MMcmd	Measurings $\widehat{H}_{ij}^{(*)} kcal/m^3$	Estimates $\widehat{H}_{ij}^{(1)} kcal/m^3$	Estimates $\widehat{H}_{ij}^{(2)} kcal/m^3$	Estimates $\widehat{H}_{ij}^{(3)} kcal/m^3$	Estimates $\hat{r}_{ij}$
1	2	3	4	5	6	7
(1,6)	170	8840	8862	8840	8840	0,043
(2,6)	90	8773	8773	8773	8773	0,030
(3,5)	110	9111	9072	9063	9111	0,074
(4,5)	180	8976	8893	8897	8976	0,047
(5,6)	70	8840	8893	8871	8840	0,047
(6,8)	140	No measuring	8805	8720	8705	0,035
(6,10)	190	8908	8873	8908	8908	0,045
(5,7)	110	No measuring	8893	8900	9097	0,047
(5,7)	110	9077	9072	9077	9077	0,074
(7,8)	155	8942	9003	9010	9097	0,064
(8,9)	295	No measuring	8909	8872	8911	0,050
(7,10)	65	8908	8934	8937	9062	0,053
(10,11)	255	No measuring	8889	8915	8947	0,047
Objective function value			17350	14971	47863	

condition (5), the criterion value drops significantly (14%).

The use of the heuristic algorithm of local optimization (solution $\widehat{H}_{ij}^{(3)}$) demonstrates its ineffectiveness. Indeed, the criterion of MLE, which characterizes the extent to which the interdependence of the measured parameters is taken into account, is significantly greater (2.75 times) than for the (solution $\widehat{H}_{ij}^{(1)}$). This proves that the heuristic algorithm can hardly be considered a satisfactory approximation to a reasonable (solution $\widehat{H}_{ij}^{(1)}$), despite its seeming natural algorithmic constructions.

XI. Conclusions

The proposed technique allows calculating the distribution of the composition of natural gas when transported via gas transmission systems of an arbitrary configuration. The developed mathematical model takes into account the irreversible and non-equilibrium nature of gas mixing processes due to various gas composition, as well as the random nature (instrument errors) of the initial data on measurements of the concentration of the components.

Instead of components concentration, formalization can be carried out in terms of gas calorific value. Using the maximum likelihood estimation, the search for the concentration distribution is reduced to the problem of mathematical programming with equality and inequality constraints, which follow from the law of mass conservation and the conditions of the process non-equilibrium. The problem is solved by well-known numerical methods using standard software packages.

The calculation example demonstrates the efficiency of the method and the potential difficulties in solving practical problems due to insufficient information and systematic instrument errors. The proposed technique can be applied to control calorific value when supplying gas to domestic consumers. It is also suitable for calculating the dew point temperature during flows through transmission systems of arbitrary configuration, which is necessary for the operational control of export supplies.

The analysis of the multicomponent gas mixture flows can serve as a methodological basis for accomplishing the prospective flow management tasks in the UGSS of the Russian Federation, which are related to the construction of gas chemical facilities using natural gas as a feedstock.

References

[1] State Standard (GOST) R 53763-2009. Combustible natural gases. Determination of dew point temperature by water, (in Russian).

[2] ISO 18453. Natural gas — Correlation between water content and water dew point. International standard First edition 2004-07-01.

[3] State Standard (GOST) R 57851.1-17 Gas condensate mixture. Part 1. Separation gas. Determination of component composition by gas chromatography. - M.: , 2017 .p. 53. (in Russian)

[4] ISO 39254.2016. Petroleum products – Determination of boiling range distribution – Gas chromatography method. p. 25.

[5] Mokhatab S., Poe W.A., Mak J. Y. *Handbook of Natural Gas Transmission and Processing Principles and Practices.* 2019.

[6] Kidnay A. J., Parrish W. R. Fundamentals of Natural Gas Processing. *Taylor & Frensis Group.* p. 433, 2006.

[7] Hébrard J. Safety of hydrogen/natural gas mixtures by pipelines: *ANR french project HYDROMEL*, Hébrard J., Studer E., Jallais S., Blanchetière V., Gentilhomme O.

[8] Kuczy´nski S. Thermodynamic and Technical Issues of Hydrogen and Methane-Hydrogen Mixtures Pipeline Transmission, Kuczy´nski S., Łaciak M., Olijnyk A., Szurlej A.,Włodek T. Energies 12(3):569 February 2019.

[9] Jieying L. Research on mixing law of natural gas

pipeline replaced by nitrogen, Jieying L., Meijuan Dang, Rui L., Xianbo G., Juntao Y. Advances in *Mechanical Engineering*, Vol. 9(6), pp. 1–8., 2017.

[10] J.W.Gibbs. Thermodynamics. Statistical mechanics. M.: *Nauka*, p.584, 1982, (in Russian.)

[11] S.R. de Groot. Thermodynamics of irreversible processes, p.280, 1956, (in Russian.)

[12] S.R. de Groot and P. Mazur, Non-equilibrium Thermodynamics, North-Holland, Amsterdam, 1962. Translation S.R. de Groot and P. Mazur, *Non-equilibrium Thermodynamics*. M.: Mir, p.456, 1964. (in Russian)

[13] Gurov K.P. Phenomenological thermodynamics of irreversible processes, M.: *Nauka*, p.127, 1978, (in Russian)

[14] Balesku R. Equilibrium and non-equilibrium statistical mechanics, p.756, 1975.

[15] Porshakov B.P., Romanov B.A. Fundamentals of thermodynamics and heat engineering. - M.: *Nedra*, 1988.

[16] Nikolaev G.P., Loyko A.E. Technical Thermodynamics. - Yekaterinburg: UrFU, p.227 2013.

[17] Purtov P. A. Introduction to non-equilibrium chemical thermodynamics. Novosibirsk: NSU, 2000.

[18] Baykov V.I. Thermophysics. Non-equilibrium heat and mass transfer processes Baykov V.I., Pavlyukevich N.V., Fedotov A.K. Shnip A.I. *Higher School*. p.476, 2018. (in Russian)

[19] Nagnibeda E., · Kustova E. Non-Equilibrium Reacting Gas Flows. Kinetic Theory of Transport and Relaxation Processes. *Springer-Verlag*, p.252, 2009. (in Russian)

[20] Berzh K. Graph theory and its applications, p.320, 1962, (in Russian)

[21] Christofides N. Graph theory: an Algorithmic Approach, p.400, 1975.

[22] State Standard (GOST) 31369-2008. *Natural gas*. Calculation of calorific value, density, relative density and Wobbe number based on component composition. (in Russian)

Nikolay Anatol'evich Kislenko, Ph.D. in Engineering. General Director of NIIgazekonomika LLC. Education: D. Mendeleev University of Chemical Technology of Russia, and St. Petersburg International Institute of Management (MBA program).

Mikhail Grigor'evich Sukharev, Sc.D. in Engineering, Professor of Gubkin Russian State University of Oil and Gas (National Research University), Moscow, Russian Federation, Chief Researcher of NIIgazekonomika LLC

Alexander Solomonovich Kazak, Sc.D. in Engineering, Professor, First Deputy General Director for Science at NIIgazekonomika LLC. Education: Gubkin Russian State University of Oil and Gas (National Research University). Specialist in the field of systems analysis and mathematical modeling.

Elena Vladimirovna Fomina received the bachelor's degree from Gubkin Russian State University of Oil and Gas (National Research University) in 2018. Now she is an engineer at NIIgazekonomika LLC.