Introduction

The concept of the stage and the criteria for selection oil fields development stages were formulated in the mid 70s of last century. The paper[1] recommends dividing oil production dynamics into four successive stages and presents some approaches to identification the boundaries between the development stages. However, these approaches are based on heuristic evaluations with no well-defined criteria. To solve this problem, we proposed a stage wise structuring based on the identification of the current oil production distribution in the individual stages and isolation of adjacent stages’ boundary points defining the duration of each stage. Using statistical analysis of experimental data on development of a large number (more than twenty) of specific fields reduced into one “generalized deposit” followed The Savitzky-Golay smoothing filter shows that the life cycle of the field can be divided into four successive stages which are definitely described as logarithmically normal, exponential, Pareto and Weibull distributions. Using non-linear logistic model for the cumulative oil production in the last fourth stage of the life cycle a computational procedure has been developed for assessing the initial recoverable reserves, as well as the method for estimating the maximum level of oil production has been proposed. The method is based on finding inflection point of the curve describing the dynamics of cumulative oil production, by examining second order differences. Based on the values of cumulative oil production in the later points of the final development stages oil production was predicted and the design parameters for the current development system were evaluated using adaptive Kalman filter in discrete time.

Keywords
Statistical modeling, Field life cycle, Stage structuring, Stage boundary, Prediction, Recoverable reserve

Furthermore, Kolmogorov-Erofeev’s equation used in the work[3] comes to the two-parameter Weibull distribution, which is only applied in case when the shift parameter δ (or “minimal run” in the terminology of the theory of reliability) is equal to zero. In general, Weibull distribution is a three-parameter one with δ ≥ 0 defined on a given sample. The third stage, when there appears water cut resulted in oil production decline, should be well described by a Pareto distribution, having a heavy right tail and reflecting hyperbolic decline in oil production.

Problems of decline curves analysis, discussed in the papers[4], apply only to the III and IV stages of the cycle. Empirical equations proposed in the above-mentioned works give the typical decline curves only in case of unique solvability. Otherwise, the question remains undecided.

We should note the variety of solving partitioning field development life cycle into the individual stages as well as the variety of their titles which exists in the literature. For example[5], consider four stages (phases) of development: developing, plateau, decline and mature and using dynamic counterparts of previous developments define the basic characteristics of field.
exploitation: the cumulative oil production, the ratio of produced oil and water, etc., and find prediction estimates of oil production and the remaining recoverable reserves. Divide the process of hydrocarbon deposits development into three phases: build-up, plateau and decline, and the fourth phase - decommissioning is still considered technically feasible for operation, though unprofitable from a financial point of view, and therefore recommend transferring the field to reserve in this phase.

By the way, the life cycle of any commercial product manufacture is also divided into four successive phases. In the paper, these stages are named as: introduction, growth, maturity and decline, like in the work: introduction to market, growth, maturity and decline.

In this paper, based on statistical study the life cycle is divided into four sequential steps, conventionally named as: build-up, developing, decline and maturity. We see that the different stages of the oil fields development differ in the type of distribution of current oil production. On the basis of maximum compactness we have developed a methodical approach to the determination of the boundary points of the adjacent stages within oil production curve accuracy which help to calculate the duration of each step. The reliability of the results is determined by the compliance rate between the theoretical and empirical distribution functions.

The numerical calculation procedure of the recoverable oil reserves and the time of peak oil production have been proposed. Moreover, the possibility to achieve design performance under the current oil production in the fourth development stage is determined based on the learning sample composed of X values at t ∈ Tj.

1.1. Data Pre-processing

We have collected the data on 20 fields from the different regions in order to investigate the life cycle of oil field development.

Let Tj = [tj, tj] denote the life cycle of field Mj (j = 1, …, j0) and let {qjt}j∈[tj,tj] denote the time series of annual oil production values in the field Mj, i.e. the given series of observations q on the discrete (with the discreteness interval Δt = 1) set of points Tj ∩ Tj.

Let us normalize the series {qjt}j∈[tj,tj] by the largest value

qjt max = maxj∈[tj,tj] qjt

So we will obtain the time series {qjt}j∈[tj,tj] for each j-th field with the values qjt from the interval [0,1]. In order to align these time series to one time interval let us normalize them in time substituting

i = tj / tj (j = 1, …, j0)

where tj is tj value, at which we can achieve the maximum value q of the factor qjt, that is tj = arg maxj qjt. As a result, under every fixed j we will obtain the time

Fig. 1 Dividing the Life Cycle into Intervals $\tilde{T}_j (j = 1, …, 4)$
series \( \{ \tilde{q}_j^i \} \), where \( \tilde{T}_j^\delta \) reflects \( T_j^\delta \) after the mentioned time substitution. With that the maximum value of these series equal to \( \max_{i \in T_j^\delta} \tilde{q}_j^i \), will be obtained in the point \( \tilde{t}_j^i = 1 \) for all \( j = 1, \ldots, J \).

Figure 2 shows the field of points corresponding to the values of time series \( \{ \tilde{q}_j^i \} \) \( j = 1, \ldots, J \). Let us denote \( i^j = \min_{j \in T_j^\delta} \tilde{t}_j^i \), \( i^\delta = \max_{j \in T_j^\delta} \tilde{t}_j^i \), \( \tilde{T}^\delta = \tilde{T}^\delta \).

The learning samples \( \{ \tilde{t}_i, \tilde{q}_i \} \) \( \in \mathbb{R}^2 \) of smoothed oil production values \( q \) in the generalized field \( M \). With that case the values \( t \) are determined as the left ends of sliding given data review window. To simplify the calculations from the above subsequence we will distinguish a partial sequence \( \{ \tilde{t}_{ik}, \tilde{q}_{ik} \} \) \( k \in \mathbb{Z} \), where \( i_k = 1 + 5(k - 1), k = 0, \ldots, 5k_0 \).

Let us divide the subsequence \( \{ \tilde{t}_{ik}, \tilde{q}_{ik} \} \) \( k \in \mathbb{Z} \) into 4 crossing partial series \( \tilde{T}_{1, \delta}^\delta, \ldots, \tilde{T}_{4, \delta}^\delta \) that fulfill the condition (1), supposing that the life cycle \( T = [t_1, t_2] \), where \( t_1 = \min_{k \in \mathbb{Z}} t_k, \ t_2 = \max_{k \in \mathbb{Z}} t_k \). It is obvious that

\[
\tilde{T}_{k, \text{learn}}^\text{learn} = \bigcup_{j=1}^{4} \{ \tilde{t}_{j, \text{learn}}^\delta \} \text{ is a discrete subset from } \tilde{T}.
\]

Let \( X_{\text{learn}}^j \) denote the set of values \( q \) corresponding \( t \) from \( \tilde{T}_{j, \text{learn}}^\delta \). The learning samples \( X_{\text{learn}}^j \) will be used below to identify the distribution of random value \( X \) (annual oil production in \( M \) field) in

\[
X_j = (x_j^l, x_j^r), \quad \text{where } x_j^l = \min_{q \in X_{\text{learn}}^j} q, \ x_j^r = \max_{q \in X_{\text{learn}}^j} q ; \ X_{\text{learn}}^j \subset X_j.
\]
1.2. Numerical Implementation of the Method

Let us divide the whole life cycle $T = [0.027; 14]$ into the learning samples $\hat{T}_{\text{learn}} = [\hat{t}_1; \hat{t}_2; \ldots; \hat{t}_n]$ where $\hat{t}_1 = 0.027$, $\hat{t}_1' = 0.648; \hat{t}_2' = 0.544, \hat{t}_2 = 1.131; \hat{t}_3 = 1.062, \hat{t}_3' = 4.477; \hat{t}_4 = 4.063, \hat{t}_4' = 14$. Learning samples $X_{\text{learn}}$ of $X$ values conform to intervals $\hat{T}_{\text{learn}}$.

1.2.1. Stage-to-stage Structuring

(a) Estimation of distribution parameters

Let us substitute $z = \lg x$ for the values $x$ of $X_i$ and advance a hypothesis that the random value $X_i$ is distributed on $X_i$ by logarithmically normal law with the distribution function $F_0(x) = \Phi\left(\frac{\lg x - \lg x_0}{\sigma_z}\right)$ where $\Phi(z)$ is distribution function of standard normal random value (with zero mean and single dispersion).

With regard to specified substitution of a variable we obtain: $\lg x_0 = z_0 = \bar{z}, \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i, z_i = \lg x_i, (n = 19)$, $\sigma = \sigma_z = s_z, s_z = \left\{\frac{1}{n-1} \sum_{i=1}^{n} (z_i - \bar{z})^2\right\}^{1/2}$.

If we insert the values of $x_i$ from $X_{\text{learn}}$ in these formulas we find $z_0 = -0.740, \sigma_z = 0.483$.

If the suggested hypothesis is true, then $X$ is distributed in $X_i$ by logarithmically normal law with the distribution function $F_0(x) = \Phi\left(\frac{\lg x + 0.74053}{0.483}\right), x > 0$.

(b) Goodness-of-fit test

At each stage of development the goodness-of-fit test of the selected hypothetical distribution function $F_0(x)$ versus the truth distribution function $F(x)$, i.e. proof of hypothesis $H_0: F(x) = F_0(x)$, will be realized owing to Kolmogorov goodness-of-fit test.

Let $x_1, \ldots, x_n$ is the sample consisting of $n$ independent observations where $n_1 (n_1 \leq n)$ values differ. Ordering them increasingly we obtain a set of variate values $x_{(1)} < x_{(2)} < \ldots < x_{(k)} < \ldots < x_{(n)}$.

Let the value $x_{(i)}$ is $k_i$ times repeated in the original sample $x_1, \ldots, x_n$. Let us construct the empirical distribution function $\hat{F}_n(x)$:

\[
\hat{F}_n(x) = \begin{cases} 
0, & -\infty < x < x_{(1)} \\
(i + k_i - 1)/n, & x_{(i)} \leq x < x_{(i+1)}, i = 1, \ldots, n_1 - 1 \\
1, & x_{(n)} \leq x < \infty
\end{cases}
\]

Kolmogorov statistic is $D_n = \sup_x |\hat{F}_n(x) - F_0(x)|$.

We construct a variation series of $x$ values of in $X_{\text{learn}}$ in order to check the hypothesis on logarithmically normal distribution of a random value $X$ on the set $X_1 = [0.019; 0.448]$, and we calculate the value $D_{ij} = \max_{x \in \hat{T}_{\text{Learn}}} |\hat{F}_{ij}(x) - F_0(x)| = |\hat{F}_{ij}(x_{(i)}) - F_0(x_{(i)})|$, at each interval $\Delta_{ij} = \{x : x_{(i)} \leq x < x_{(i+1)}\}$ where $\hat{F}_{ij}(x)$ is calculated by Eq. (2). The calculated value $D_n$ of $D_n$ statistic is found of the relationship

\[
D_n = \max_{i=1}^{n} D_{ij}.
\]

For $X_{\text{learn}}$ with $n = 19$ by Eq. (3) we will obtain $D_{ij} = 0.18$. The critical value for Kolmogorov criterion of goodness if $n = 19$ and $\alpha = 0.20$, is equal to 0.237. As $D_n < 0.237$, the hypothesis $H_0: F(x) = F_0(x)$ if $x \in X_1$, can be accepted with the confidence probability $P_1 = 0.8$.

Similarly, the distribution functions are identified for the remaining stages. The calculation results of distribution functions and confidence probability of fitting criterion for all four stages are shown in Table 1. These confidence probabilities can be improved by reducing the observation scale.

We should note at each stage of the identification of the real distribution function $F(x)$ several competing distributions were considered as a theoretical distribution function $F_0(x)$: normal, logarithmically normal, exponential, Pareto and Weibull distribution. Preference was given to the distribution law, when with the greatest confidence level could be accepted the hypothesis $H: F(t) = F_0(t)$ by the Kolmogorov goodness-of-fit criterion.

So, the entire oil field life cycle can be divided into four sequential stages: build-up, developing, decline and maturity, which are reliably (80 % at least), described to be respective logarithmically normal, exponential, Pareto and Weibull distributions.

The first period (build-up) is the initial stage of development with the unsteady regime and the adaptive selection of production engineering in accordance with the filtration properties of the reservoir. The second period (developing) is associated with an increase of percent of oil recovery from the reservoir and continues until the maximum level of oil production and the possible stabilization of this level as saturation level (the so-called "plateau"). The third period (decline) exhibits a sharp decline in oil production due to flooding and formation contamination. Finally, the fourth period (maturity) reflects the comprehensive application of natural operation methods with a fluctuating relative to some minimum level of oil production, which requires additional (artificial) methods of enhanced oil recovery.

1.2.2. Determination of Adjacent Stages’ Limits

In accordance with boundary detection method between adjacent stages, described in Appendix A, we check the intersection possibility of graphs $f_1(x)$ and $f_2(x)$, $f_3(x)$ and $f_4(x)$, $f_5(x)$ and $f_6(x)$, $f_7(x)$ and $f_8(x)$, where $f_i(x)$, $f_3(x)$, $f_4(x)$ and $f_8(x)$ are the probability density functions corresponding to stages I, II, III and IV.
Table 1 The Calculation Results of Distribution Functions and Confidence Probability of Fitting Criterion for All Four Stages

| Stages | Distribution law | Distribution function | Probability belief of fitting criterion |
|--------|------------------|-----------------------|----------------------------------------|
| I      | Logarithmically normal $F_1(x) = \Phi \left( \frac{\ln x + 0.740}{0.483} \right)$, $x > 0$ | $P_1 = 0.8$ |
| II     | Exponential $F_2(x) = \begin{cases} 1 - \exp \left( -4.861(x - 0.388) \right), & x \geq 0.388 \\ 0, & x < 0.388 \end{cases}$ | $P_2 = 0.8$ |
| III    | Pareto $F_3(x) = \begin{cases} 1 - \left( \frac{0.254}{x} \right)^{1.181}, & x \geq 0.254 \\ 0, & x < 0.254 \end{cases}$ | $P_3 = 0.99$ |
| IV     | Weibull $F_4(x) = \begin{cases} 1 - \exp \left( -\frac{(x - 0.003)^{1.181}}{0.011} \right), & x \geq 0.003 \\ 0, & x < 0.003 \end{cases}$ | $P_4 = 0.85$ |

According to the above method of detection of stages neighbouring boundaries we check the possibility of $f_i(x)$ and $f_j(x)$, $f_k(x)$ and $f_l(x)$ plots crossing.

We have $A = 0.027$ and $B = 14$ for the field $M$. The analysis of current oil production frequency curves showed that $f_i(x)$ disjoints $f_j(x)$ in the set $G_{1,2} = [0; \infty] \cap [0.388; \infty] \cap [A, B] = [0.388; 14]$; $f_2(x)$ disjoints $f_3(x)$ in the set $G_{2,3} = [0.388; \infty] \cap [0.254; \infty] \cap [A, B] = [0.388; 14]$ and finally $f_3(x)$ disjoints $f_4(x)$ in the set $G_{3,4} = [0.254; \infty] \cap [0.004; \infty] \cap [A, B] = [0.254; 14]$.

Then by Eq. (A-4) we obtain

$x_{1,2}^{\text{bound}} = \max \{0; 0.388\} = 0.388$;
$x_{2,3}^{\text{bound}} = \max \{0.388; 0.254\} = 0.388$;
$x_{3,4}^{\text{bound}} = \max \{0.254; 0.004\} = 0.254$.

As $\tilde{T}_{1,2} = [0.544; 0.648]$, $\tilde{T}_{2,3} = [1.062; 1.131]$ and $\tilde{T}_{3,4} = [4.063; 4.477]$, by Eq. (A-2) we find $t_{\text{bound}}^{1,2} = 0.544$,
$t_{\text{bound}}^{2,3} = 1.131$, $t_{\text{bound}}^{3,4} = 4.063$.

2. The Determination of Initial Recoverable Reserves and Peak Level of Current Oil Production

Let $Q(t) = \sum_{t=t_0} q(t')$ denote cumulative oil production at the moment $t$. As current oil production is relatively stable at the closing stage, there exists a finite limit $Q_\infty = \lim_{t \to \infty} Q(t)$ equal to the initial recoverable reserves. Below we present the procedures to determine the value $Q_\infty$ and the peak level $q_{\text{max}}$ of current oil production.

For calculation of value $Q_\infty$ we propose to use the linear regression $y = a_0 + a_1 x$, where $x = Q$ and $y = q/Q$, $q$ and $Q$ is a current and cumulative oil production. The coefficients $a_0$ and $a_1$ are obtained by the least-squares method (LSM) on the basis of $\{q_i, (q / Q)_i\}$ observations at the moments $t_i (i = 1, \ldots, n)$ on IV stage of the life cycle of field development. According to the LSM estimates we calculate $\tilde{a}_0$ and $\tilde{a}_1$.

$$Q_\infty = \frac{\tilde{a}_0}{\tilde{a}_1}. \quad (4)$$

Detailed proof of Eq. (4) is given in Appendix B.

For estimation of the peak level of current oil production $q(t)$ we find a critical function point $q(t)$ (let us denote it $t_{\text{peak}}$), which is equivalent to finding the flex point of autocatalytic curve $Q(t)$, where $Q(t)$ is the cumulative oil production per year.

The computational procedure for assessing the value $t_{\text{peak}}$ and the calculated values $t_{\text{peak}}$ and $q(t_{\text{peak}})$ are listed in Appendix B.

It should be noted that initial recoverable reserves $Q_\infty$ can be calculated only in the IV stage, whereas a maximum level of current oil production $q(t_{\text{peak}})$ in the II stage of lifetime of field development.

3. Predicted Oil Production

As shown above, the probabilities distribution function of values $x$ of the random value $X = q(t)$ at fourth stage of development (that is if $t \in T$) is well described by Weibull distribution

$$F(x) = 1 - \exp \left( -\frac{(x - \delta)^\beta}{\theta} \right),$$

where $\delta = 0.003$, $\beta = 1.104$, $\theta = 0.011$.

As we have the discrete values of the random variable $X$, then $F(x) = P(X \leq x) = \sum_{i:A_i \leq x} p_i$, where
\( p_t = P(x = q_t = q(\tilde{t})) \). Then \( p_t = q_t / Q_\infty \) due to geometric probability and, it follows that \( F(x_{x=q(\tilde{t})}) = Q(\tilde{t})/Q_\infty \).

Let \( \tilde{t}_0 \) be such a point from the interval \( \tilde{T}_4 = [4.063; 14] \) that the interval \( [\tilde{t}_0; 14] \) contains at least \( n (n \geq 50) \) measurements \( x(\tilde{t}) = q(\tilde{t}) \). Let formulate function \( \tilde{x}(\tilde{t}) \), obtained from \( x(t) \) following substitution \( \tilde{x}(\tilde{t}) = x(\tilde{t} + t_0) - \delta \) (\( \delta = 0.003 \) is the \( F_1(x) \) function parameter), as a scaling relationship \( \tilde{x}(\tilde{t}) = A / \tilde{t}^\mu , \mu > 0 \). Then, \( \tilde{t} \in [\tilde{t}_1, \tilde{t}_2] \), \( \tilde{t}_1 = 0, \tilde{t}_2 = 14 - \tilde{t}_0 \).

The parameters \( A \) and \( \mu \) of this relation are defined through regression \( Y = \alpha_0 + \alpha_1X \), where \( Y = \ln \tilde{x}, X = \ln \tilde{t}, \alpha_0 = \ln A, \alpha_1 = -\mu \). We estimate \( \hat{\alpha}_0 = -6.706, \hat{\alpha}_1 = -0.128 \) with the aid of LSM, whence we obtain assessed values \( \hat{A} = e^{\hat{\alpha}_0}, \hat{\mu} = -\hat{\alpha}_1 \).

After the above substitutions with regard to scaling relationship the distribution function \( F(x) \) of values \( x(\tilde{t}) = q(\tilde{t}) \) when \( \tilde{t} \in [\tilde{t}_1, \tilde{t}_2] \) is written as

\[
F_{\tilde{x}}(\tilde{t}) = 1 - \exp\left\{-b \cdot \tilde{t}^\mu \right\}, \quad b = 0.0555, \quad d = 0.128.
\]

Then

\[
\lambda(\tilde{t}) = F'_{\tilde{x}}(\tilde{t}) = a b \tilde{t}^{-\mu - 1} = 0.007 / \tilde{t}^{0.872}.
\]

The cumulative normalized oil production \( Q_{\infty}(\tilde{t}) = Q(\tilde{t}) / Q_\infty \) satisfies the equation

\[
\frac{dQ_{\infty}(\tilde{t})}{d\tilde{t}} = \lambda(\tilde{t})(1 - Q_{\infty}(\tilde{t})), \quad \tilde{t} \in [\tilde{t}_1, \tilde{t}_2] \quad (5)
\]

where \( \lambda(\tilde{t}) = ab \tilde{t}^{-\mu + 1} \) is oil reserves production rate.

The Eq. (5) is used in \(^3\) to describe oil production dynamics and is the analogue of Kolmogorov-Erofeev equation for chain branch chemical reactions. For cumulative oil production \( Q(\tilde{t}) \) from Eq. (5) we obtain the following equation

\[
\frac{dQ(\tilde{t})}{d\tilde{t}} = -\lambda(\tilde{t})Q(\tilde{t}) + \lambda(\tilde{t})Q_\infty + w(\tilde{t}) \quad (6)
\]

where \( w(\tilde{t}) \) is admissible random disturbance affecting the process state. If, moreover, the state \( x(\tilde{t}) = Q(\tilde{t}) \) is measured by the value

\[
y(\tilde{t}) = C(\tilde{t}) x(\tilde{t}) + u(\tilde{t}) \quad (7)
\]

where \( u(\tilde{t}) \) is the measurement noise, then Eqs. (6)-(7) describe the linear dynamic system in continuous time.

The key points of a suggested statistical simulation method of oil field development lifetime are the following:

(a) Partition of lifetime into four consecutive stages: build-up, developing, decline and maturity;
(b) Identifying a probability distribution of a current oil production value \( q(t) \) by virtue of standard distributions: log-normal, exponential, Pareto and Weibull distributions for I, II III and IV stages accordingly;
(c) Alternation of probability distributions laws of a value \( q(t) \) of quantity testifies transition to a subsequent stage of development;
(d) A determination method of boundary points of adjacent stages has been suggested and this method characterizes duration of a previous stage;
(e) Calculating formulas have been given for a maximum production level and initial recoverable reserves by values \( q(t) \) for II and IV stages accordingly;
(f) Forecast models have been developed for cumulative oil production \( Q(t) \). These models enable assessing a possible level of reaching designed production economics in the near future and correcting a strategy of further field development on the basis of comparison of designed and predictable recoverable reserves.

**4. Application Example**

For illustration of a suggested statistical simulation method of lifetime, we will consider an analysis example of a development condition of one specific horizon of Uzen field (Kazakhstan).

Let us consider the current oil production history from the start of “Zhetybay” field production (Fig. 4) (conventional production values have been presented).

As is clear from Fig. 4 current oil production history is visually divisible into two parts 1965-1999 and since 1999 up to date. It is due to the fact that up to 1999 a part of the field was on-stream, whereas the remaining
part of the field was drill out after 1999.

4. 1. Separation of Development Stage over a Period of 1965-1999

The calculation data of distribution function and goodness-of-fit test probability belief for all the four stages are listed in the Table 2.

Thus, the entire life cycle of the field is possible to divide onto four consecutive steps: build-up, developing, decline and maturity, which with plenty good enough accuracy are described by respectively logarithmically normal, exponential, Pareto and Weibull distributions: $T_1 = [01.1965; 01.1972]$, $T_2 = [01.1972; 10.1975]$, $T_3 = [10.1975; 01.1995]$, $T_4 = [01.1995; 12.1999]$ (Fig. 5). Reference point $t_\cdot 0$ in Fig. 5 corresponds to commencement date of development 01.1965.

4. 2. Evaluation of Initial Recoverable Reserves

Distribution function for value $X = q(t) \cdot (q(t) \cdot$ current oil production) on IV stage can be written in the form:

$$F_4(x) = 1 - \exp \left\{ -\frac{(x - \delta)^\beta}{\theta} \right\}, \quad x \geq \delta$$

where $\delta = 44.76$, $\beta = 2.67$, $\theta = 134.925$.

Subsequent to substitution $x(t) = x(t_0) - \delta$, $t_0 = 29$ (which is equivalent of date 01.1995) let us represent function $x(t)$ as a scaling relationship $\tilde{x}(t) = \lambda(t) t^\mu$.

Let us construct a regresional relationship $Y = \ln \tilde{x}$ from $X = \ln(\tilde{x} - t - t_0)$ by points $t \in T_4 = [01.1995; 12.1999]$ (which in order provisional scale corresponds to interval $[29, 34]$)

$$Y = \alpha_0 + \alpha_1 X.$$ 

LS method estimate of parameters $\alpha_0$, $\alpha_1$:

$\hat{\alpha}_0 = 4.02$, $\hat{\alpha}_1 = 0.05$.

Whereof we obtain estimate of parameters $\lambda$ and $\mu$:

$$\hat{\lambda} = e^{\hat{\alpha}_0} = 55.496, \quad \hat{\mu} = -\alpha_1 = -0.05.$$

Subsequent to above substitutions, with regard to scaling relationship, we obtain

$$F_4(\tilde{x}) = 1 - \exp \left\{ -b\tilde{t}^\mu \right\}, \quad b = 48.122, \quad a = 0.036.$$

Then

$$F_4(\tilde{t}) = ab\tilde{t}^{\mu-1} \exp \left\{ -b\tilde{t}^\mu \right\}.$$

Cumulative normalized oil production $Q_{\alpha}(\tilde{t}) = Q(\tilde{t})/Q_{\alpha}$ meets the difference equation

$$\frac{dQ_{\alpha}(\tilde{t})}{d\tilde{t}} = \tilde{\lambda}(\tilde{t}) (1 - Q_{\alpha}(\tilde{t})), \quad \tilde{t} \in [\tilde{t}_1, \tilde{t}_2]$$

where $\tilde{\lambda}(\tilde{t}) = ab\tilde{t}^{\mu-1}$ is oil reserves production rate:

$$\tilde{t}_1 = 0, \quad \tilde{t}_2 = 34 - 29 = 5.$$

Going to discrete values (in increments of discreteness $\tau$) $x_k = Q(\tilde{t}_k)$, $x_k = \tilde{\lambda}(\tilde{t}_k)$, $A_k = (1 - \tilde{\lambda}_x)$, $U_x = \tau x_k Q_{\infty}$ from differential Eq. (5) we obtain difference equation system:

![Fig. 5 Separation of Field Development Stages over a Period of 1965-1999](image-url)
where $w_k$ and $v_k$ are noise process status and experimental observation troubles respectively.

The first of the simultaneous equations is an equation of state, the second one is an experimental observation equation (in the calculations has been accepted $C = 1$).

Using the Kalman algorithm of optimal linear filtering with discrete time periods we obtain the forecasts $Q(t)$ in the period [1965-1999] and [2000-2011].

As is clear from Fig. 6 when forecast in the period [1965-1999] the initial reserves composed 72.5 mln tones, i.e. total 63 % of rated initial recoverable reserves (3D geological field model based estimated original oil in place), whereas subsequent to drilling-out operation of the deposit since 2000 the initial recoverable reserves under the forecast in the period [2000-2011] composed 106.8 mln tones, which makes 94 % of rated initial recoverable reserves.

Further stabilization, oil production enhancement and cumulative gain in recoverable reserves in the field require application of state-of-the-art bottom-hole zone stimulation methods. Hence, the proposed approach enables to evaluate the efficiency of the strategy used to plan and develop activities on enhancement the level of efficiency achieved.

Conclusions

(1) Based on the identification of current oil production values distribution at the separate stages we propose the statistic method to detect boundary points of the neighbouring stages and determine the durability of each stage of the oil field development. With that we show that the distribution of current oil production values at the consecutive stages I-IV with the sufficiently high confidence is approximated according to lognormal, exponential, Pareto and Weibull distributions.

(2) Availability of every consecutive stage duration enables defining a maximum level of current oil production in the II stage and calculating initial recoverable oil reserves in the IV stage under a current field development strategy.

(3) We have shown that Weibull distribution function of current oil production values at the fourth stage of the life cycle development, relying on the $q$-$t$ scaling dependence, aligns with the linear differential equation solution relatively to the normalized cumulative oil production that is the analogue of Kolmogorov-Erofeyev kinetic equation.

(4) With the help of Kalman adaptive filter we have obtained the calculation formulas for the one-step and multistep oil production prediction at the closing stage of development with the estimation of the prediction error and its variance.

Nomenclatures

\[ \begin{align*}
    x_{k+1} &= A_k x_k + U_k + \tau w_k \\
    y_k &= C_k x_k + v_k
\end{align*} \]

where $w_k$ and $v_k$ are noise process status and experimental observation troubles respectively.

The first of the simultaneous equations is an equation of state, the second one is an experimental observation equation (in the calculations has been accepted $C = 1$).

Using the Kalman algorithm of optimal linear filtering with discrete time periods we obtain the forecasts $Q(t)$ in the period [1965-1999] and [2000-2011].

As is clear from Fig. 6 when forecast in the period [1965-1999] the initial reserves composed 72.5 mln tones, i.e. total 63 % of rated initial recoverable reserves (3D geological field model based estimated original oil in place), whereas subsequent to drilling-out operation of the deposit since 2000 the initial recoverable reserves under the forecast in the period [2000-2011] composed 106.8 mln tones, which makes 94 % of rated initial recoverable reserves.

Further stabilization, oil production enhancement and cumulative gain in recoverable reserves in the field require application of state-of-the-art bottom-hole zone stimulation methods. Hence, the proposed approach enables to evaluate the efficiency of the strategy used to plan and develop activities on enhancement the level of efficiency achieved.

Conclusions

(1) Based on the identification of current oil production values distribution at the separate stages we propose the statistic method to detect boundary points of the neighbouring stages and determine the durability of each stage of the oil field development. With that we show that the distribution of current oil production values at the consecutive stages I-IV with the sufficiently high confidence is approximated according to lognormal, exponential, Pareto and Weibull distributions.

(2) Availability of every consecutive stage duration enables defining a maximum level of current oil production in the II stage and calculating initial recoverable oil reserves in the IV stage under a current field development strategy.

(3) We have shown that Weibull distribution function of current oil production values at the fourth stage of the life cycle development, relying on the $q$-$t$ scaling dependence, aligns with the linear differential equation solution relatively to the normalized cumulative oil production that is the analogue of Kolmogorov-Erofeyev kinetic equation.

(4) With the help of Kalman adaptive filter we have obtained the calculation formulas for the one-step and multistep oil production prediction at the closing stage of development with the estimation of the prediction error and its variance.

Nomenclatures

\[ \begin{align*}
    \text{A} & \text{h} & \text{B} \text{: entire life cycle of the field} \\
    A_j & \text{: } F_j(x) \text{ support} \\
    c_j & \text{ and } d_j \text{ : starting and end point of interval } A_j \\
    D_j^{(x)} & \text{: Kolmogorov distance between } F_j(x) \text{ and } F_j(x) \\
    D_\kappa & \text{: Kolmogorov statistic} \\
    F_j(x) \text{ and } f_j(x) & \text{: integral and differential functions of current oil production distribution on stage } j \\
    \hat{F}_j(x) & \text{: empiric distribution function} \\
    G_{ij} & \text{: meet of sets } A_i, A_j \text{ and } [A, B] \\
    h_{\text{up}} & \text{: Weight coefficients of The Savitzky-Golay filter} \\
    \hat{k}_{ij} & \text{: estimate of goodness of fit probability } F_j(x) \text{ and } F_j(x) \text{ at } G_{ij} \\
    P_j & \text{: confidence probability of decision } H: F_j(x) = F_0(x) \\
    P^*_j(t) & \text{: The values of Gram polynomial of degree } k \text{ in points } 2m + 1 \\
    q(t) & \text{: current oil production per annum } t \\
    Q(t) & \text{: cumulative oil production per annum } t \\
    \omega & \text{: residual recoverable petroleum reserves under the existing deposit development strategy} \\
    \text{peak} & \text{: peak time } t_{\text{peak}} \text{ of current oil production} \\
    t_{\text{bound}} & \text{: phase boundary } j \text{ and } j \text{ of the field life cycle} \\
    \tilde{T} & \text{: subset of } [A, B], \text{ upon which we identify accumulated distribution of current oil production on stage } j \\
    \tilde{t}_j & \text{ and } \tilde{T}_j \text{: starting and end point of interval } \tilde{T}_j \\
    \tilde{T}_{j, \text{ } \text{cross point}} & \text{: meet of sets } \tilde{T}_j \text{ and } \tilde{T}_j \\
    \tilde{t}_{\text{cross point}} & \text{ and } \tilde{T}_{j, \text{ } \text{cross point}} \text{: starting and end point of interval } \tilde{T}_{j, \text{ } \text{cross point}} \\
    \text{X}_{\text{cross point}} & \text{: oil production at the moment } t_{\text{cross point}} \\
    X_{\text{cross point}} & \text{: teaching selection of current oil production values in the period } \tilde{T}_j \\
    x_{\text{cross point}} & \text{ with } f_j(x) \text{ cross point} \\
    v(t) & \text{: constitutive equations perturbation} \\
    \delta, \theta, \beta & \text{: Weibull parameter} \\
    \text{tr}(t) & \text{: measuring noise}
\end{align*} \]

References

1) Ivanova, M., Semin, E. I., Surguchev, M. L., Baishev, B. T., 10th World Petroleum Congress, Bucharest, September, 1979. 
2) Bocharov, V. A., Grigoriev, M. M., J. Oil Industry, 1, 24 (2002).
Appendix A. Detection of Adjacent Stage Limits

Let the random value \( X = q(t) \) at \( t \in T \) belong to the interval \([A, B]\) and \( t \in \hat{T}_j \) at be distributed in \( X_j \subset [A, B] \) with the continuous distribution function \( F_j(x) \), having a support (that is set of \( x \) points, where, except null set, \( F_j(x) \neq 0 \) \( A_j \in \{c_j, d_j\} \) and having density \( f_j(x) = F_j'(x) \).

Definition. Let us call the point \( x^0_{\text{bound}} \) a limiting point of distribution functions \( F_j(x) \) and \( F_j^+(x) \), corresponding to the change of \( X = V(t) \) value at the two adjacent stages \( T_j \) and \( T_{j'} \), if at \( x = x^0_{\text{bound}} \) we observe a continuous transition or a snap-back of distribution function of \( X \) value from \( F_j(x) \) to \( F_j^+(x) \) function, i.e.:

\[
F(x) = \begin{cases} 
F_j(x), & x < x^0_{\text{bound}}^j \\
F_j^+(x), & x \geq x^0_{\text{bound}}^j.
\end{cases}
\]  

(A-1)

The assessed value \( \hat{i}^0_{\text{bound}} \) of the boundary \( i^0_{\text{bound}} \) of adjacent stages \( T_j \) and \( T_{j'} \) is calculated by Eq. (A-2):

\[
\hat{i}^0_{\text{bound}} = \arg \min_{i \in [1, j]} \left| x^0_{i, j} - q(t) \right|
\]  

(A-2)

We propose the following calculation procedure to find the value \( x^0_{\text{bound}} \).

Let us consider two cases:

1. \( f_j(x) \) and \( f_{j'}(x) \) disjoint in the set \( G_{j,j'} = A_j \cap A_{j'} \cap [A, B] \); and
2. \( f_j(x) \) and \( f_{j'}(x) \) intercross in the set \( G_{j,j'} \) in the points \( x^0_{i,j} < x^0_{i,j'} < \ldots < x^0_{k,j'} \) where \( R_{j,j'} \geq 1 \) is the total number of \( x^0_{i,j'} \) points.

In the first case

\[
x^0_{\text{bound}} = \max\{c_j, c_{j'}\}.
\]  

(A-3)

In the second case we will use the maximum compaction method\(^{56}\). Following\(^{56}\), let \( D^0_{1,j} \) denote the Kolmogorov distance between the distribution functions \( F_j(x) \) and \( F_{j'}(x) \) in the point \( x^0_{i,j'} \):

\[
D^0_{1,j} = F_j\left(x^0_{i,j'}\right) - F_{j'}\left(x^0_{i,j'}\right).
\]

Within the maximum compaction method the probability of goodness of fit in the set \( A_{j,j'} = A_j \cap A_{j'} \) of continuous distribution functions \( F_j(x) \) and \( F_{j'}(x) \) will be determined through the expression \( k_{j,j'} = 1 - \sum_{r=1}^{R_{j,j'}} D^0_{1,j} \).

The expression for \( k_{j,j'} \) is simplified, if the distribution densities \( f_j(x) \) and \( f_{j'}(x) \) are similar only in one point \( x = x^0_{i,j} \) of \( A_{j,j'} \). In this special case \( k_{j,j'}(x) = 1 - D^0_{1,j} \left(x^0_{i,j}\right) \), where \( D^0_{1,j} (x) = |F_j(x) - F_{j'}(x)| \) is Kolmogorov distance between the probability distribution functions \( F_j(x) \) and \( F_{j'}(x) \).

Let us denote, \( k_{j,j'}(r) = 1 - \sum_{r=1}^{R_{j,j'}} D^0_{1,j} \), \( r \in [1, 2, ..., R_{j,j'}] \) and calculate \( x^0_{i,j} = \arg \max_{r \in [1, R_{j,j'}]} k_{j,j'}(r) \).

Relying on confidence probabilities \( P_j \) and \( P_{j'} \) to accept hypotheses \( F(x) = F_j(x) \), \( x \in X_j \), and \( F(x) = F_{j'}(x) \), \( x \in X_{j'} \), the goodness of fit for \( F_j(x) \) and \( F_{j'}(x) \) on the set \( G_{j,j'} \) is estimated as \( \hat{k}_{j,j'} = k_{j,j'} \left(x^0_{i,j}\right) \cdot P_j \cdot P_{j'} \).

The decision rule for decision making is:

\[
x^0_{\text{bound}} = \begin{cases} 
x^0_{i,j} & \text{if } k_{j,j'} \geq k^*, \\
\max\{c_j, c_{j'}\} & \text{if } k_{j,j'} < k^*.
\end{cases}
\]  

(A-4)

At \( \hat{k}_{j,j'} < k^* \) we may regard \( f_j(x) \) and \( f_{j'}(x) \) as disjoint ones and use Eq. (A-3) to calculate \( x^0_{\text{bound}} \). Evidently in case, when high limit of current oil production \( q_{\text{max}} \) is
obtained in the only point \( x_0 = q(t_{\text{peak}}) \), we naturally accept \( x_0 \) and \( t_{\text{bound}}^{2,3} = t_{\text{peak}} \) respectively as frontier point \( t_{\text{bound}}^{2,3} \).

Tests of the method of maximum compactness for identifying a Gaussian distribution, Laplace or Cauchy distribution rated \( k_{j/} \) value not exceeding 0.7[6]. Therefore, in the numerical calculations, in the Eq. (A-4) we assumed a threshold value of \( k^* \) equal to 0.7.

**Appendix B. Estimation Method of High Limit for Current Oil Production**

Let us expand \( \frac{dQ}{dt} \) function in a Maclaurin series in powers of \( Q \):

\[
\frac{dQ}{dt} = c_0 + c_1 Q + c_2 Q^2 + c_3 Q^3 + \cdots.
\]

As \( \frac{dQ}{dt} = 0 \), if \( Q = 0 \), then \( c_0 = 0 \). Where as equation
\[
\frac{dQ}{dt} = 0
\]

is also fulfilled when \( Q = Q_\infty \) that under conditions \( \frac{dQ}{dt} \) when \( Q = 0 \) and \( Q = Q_\infty \) and the smallest order of the degree \( Q \) being a part of expansion of the function is equal to two we obtain the equation

\[
\frac{dQ}{dt} = c_1 Q + c_2 Q^2.
\]

Assuming \( a = c_1 \) and \( -b = c_2 \), let us write this equation as a nonlinear logistic model which was first suggested by R. Ferhulst in 1929:

\[
\frac{dQ}{dt} = aQ - bQ^2 \tag{B-1}
\]

where \( a \) is a growth coefficient of oil withdrawal; \( b \) is a loss coefficient of oil withdrawal; \( a > 0, b > 0 \).

Dividing the both parts of the Eq. (B-1) by \( Q \), we will obtain

\[
\left( \frac{dQ}{dt} \right)/Q = a - (a/Q_\infty) Q. \tag{B-2}
\]

This is the straight line equation with the angular coefficient \(-a/Q_\infty\), which crosses the vertical axis at \( \left( \frac{dQ}{dt} \right)/Q = a \) and the horizontal axis at \( Q = Q_\infty \).

Therefore, assuming that the dependence \( \left( \frac{dQ}{dt} \right)/Q \) on \( Q \) satisfies the Eq. (B-2), we obtain the following simple procedure of finding the constant \( Q_\infty \).

By the sample \( \left\{ Q_i, \left( \frac{q}{Q_i} \right) \right\} \) of observations at the moments \( \{\dot{t}_i\}, i = 1, \ldots, n \), at stage IV of the deposit development life cycle we construct the linear regression

\[ y = \alpha_0 + \alpha x, \] \( x = Q \) and \( y = q/Q \). This straight line will cross the axis \( y \) at \( q/Q = \alpha_0 \) and the axis \( x \) at \( Q = -\alpha/\alpha_0 \). From where we find \( Q_\infty = -\alpha/\alpha_0 \). In the calculations for the generalized field \( M \) we obtained \( Q_\infty = 35.8 \).

With the aim of estimation the maximum current oil production level \( q(t) \) (let us denote it \( t_{\text{peak}} \)) the condition \( \frac{dQ}{dt} = 0 \) is required, which is equivalent to the equation

\[
\frac{d^2Q}{dt^2} = 0, \text{ i.e. a necessity of the autocatalytic curve bend} \ Q(t) \text{ in the point} \ t = t_{\text{peak}} \text{ determined with the Eq. (B-2); in addition} \ Q(t_{\text{peak}}) = Q_\infty / 2^{17}. \]

It means it is sufficient to find the inflection point \( t_q \) of functions \( Q(t) \) as \( t_q = t_{\text{peak}} \) in order to determine \( t_{\text{peak}} \). Let us denote \( Q(t) = X_n \) when \( t = t_q \). As \( \frac{d^2Q}{dt^2} \) function changes sign from \( + \) to \( - \) when crossing \( t_q \) point in the order of increasing \( t \), then the second difference \( \Delta^2 X_n = X_{n+1} - 2X_n + X_{n-1} \) should have the same property when the crossing \( n = n^* \) point in the order of increasing \( n \), where \( X_{n^*} = Q(t_q) \). Relying on the given case we may assume the following numerical procedure to search \( t_{\text{peak}} \).

Starting with the \( n = n_0^* \) point, which corresponds to the II stage start-up, we will calculate \( \Delta^2 X_n \) and compare this value with a certain fixed threshold \( \varepsilon (\varepsilon > 0 \text{ is a sufficiently small number}) \). To satisfy the in equation \( \Delta^2 X_n < \varepsilon \), we let take the point \( n \) as the preliminary estimate of \( n^* \) of the function flex point \( Q(t) \). By the values \( Q \) in the points \( \{\dot{t}_0, \ldots, \dot{t}_m\} (m \geq 10) \)

we construct the regression model \( \tilde{Q}(t) \) (nonlinear one, in general). Extrapolating \( \tilde{Q}(t) \) to the moments \( \dot{t}_0, \dot{t}_1, \ldots \), calculating the corresponding values \( \Delta^2 X_n \), we find the first point \( \tilde{n} \geq n^* + 1 \), in which \( \Delta^2 X_n \leq 0 \). Then we may accept \( \tilde{t}_{\text{peak}} = \tilde{n} - 1 \) for the estimate of \( t_{\text{peak}} \) constant.

For the field \( M \) by the values \( Q \) in the points from interval \( \tilde{t} \in [0.54454; 0.92406] \) we determine the point \( \tilde{t}_q = 0.92406 \) in which \( \Delta^2 X_n < \varepsilon \) when \( \varepsilon = 0.1 \). The best nonlinear regression constructed by the values \( Q \) in the points \( \{\dot{t}_0, \ldots, \dot{t}_m\} (i = 0, 1, \ldots, m) \) is presented as the dependence

\[
\tilde{Q}(t) = \left( a + c t^2 + d t^4 + g t^6 + h t^8 \right) / \left( 1 + b t^2 + d t^4 + f t^6 + h t^8 + j t^10 \right)
\]

with coefficients \( a = 0.097852352, b = -2.6851843, c = 11.658915, d = \ldots \)
2.6238334, $e = -32.259534, f = -1.3468213, g = 29.519634, h = 0.62538594, i = -8.8837626, j = -0.20485559$ and the coefficient of nonlinear correlation $r^2 = 1$. Calculated by the predicted values $\hat{Q}(t)$ when $t > 0.92406$ the second differences $\Delta^2 X_n \left( X_n = \hat{Q}(t_n) \right)$ change the sign from $(+)$ to $(-)$ in the point $t = 1.02756$ which is accepted as the assessed value of $\hat{t}_{\text{peak}}$. The calculated value $q(\hat{t}_{\text{peak}}) = 0.949$ aligns with the actual value $q_{\text{max}} = 0.94878$ within the accuracy to the second sign.