S©MR

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ
Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru
ISSN 1813-3304

Vol. 20, No. 2, pp. 144–145 (2023)
https://doi.org/10.33048/semi.2023.20.???

УДК 551.468
MSC 86A05

CLUSTER ANALYSIS OF LARGE-SCALE HYDRODYNAMIC CHARACTERISTICS OF THE ARCTIC SEAS BASED ON THE RESULTS OF EDDY-PERMITTING MODELING

G.A. PLATOV AND **E.N. GOLUBEVA**

Communicated by P.P. Petrov

Abstract: The characteristics of the eddy mass transport are estimated depending on the values of the parameters of a large-scale flow that forms under the conditions of the shelf seas in the Arctic. For this, the results of numerical simulation of five Arctic seas with a horizontal resolution permitting the development of mesoscale eddies are used. The parameters obtained in numerical experiments are considered as a statistical sample and are analyzed for their independence. Further clusterization in space of found set of most independent variables allows to identify several separate types of eddies. The features of the resulting clustering are described and discussed in the presented work.

Keywords: eddy mass transport, subgrid-scale processes, parametrization, clustering.

Platov, G.A., Golubeva, E.N., Cluster analysis of large-scale hydrodynamic characteristics of the Arctic seas based on the results of eddy-permitting modeling.

The work was carried out during the implementation of project No. 123081400010-2 and supported by the Ministry of Science and Higher Education of the Russian Federation (agreement number 075-03-2024-349/2). The results of numerical experiments and their analysis used were obtained within the framework of the continued RSF project No. 20-11-20112.

Received November, 1, 2023, Published December, 31, 2023.

1 Introduction

The eddy transport of scalar quantities is one of the subgrid-scale processes, the direct description of which within the framework of the oceanic block of large-scale climate models is not yet possible due to the accepted horizontal and vertical resolution, as well as due to a number of simplifying approximations. The scales of eddy formations in the Arctic ocean vary quite widely, and not all of them can be adequately described. These include mesoscale and submesoscale eddies whose sizes are determined by the Rossby deformation radius and range from several hundred meters to several kilometers. The essential exchange of water properties caused by the eddy motion, requires the search for additional possibilities of their description in large-scale models by means of parametrization.

In our previous work [1] we considered the parametrization of the eddy mass transport $-\left(\overline{\rho'\vec{u'}}\cdot\vec{n}\right)$ in the direction $\vec{n}$ of the bottom slope in the form of a horizontal diffusion flux, taking into account the counter-gradient component [2, 3, 4]

$$-\left(\overline{\rho'\vec{u'}}\cdot\vec{n}\right)=K\frac{\partial\bar{\rho}}{\partial\vec{n}}+q, \tag{1}$$

where K is the eddy diffusion coefficient, $\frac{\partial\bar{\rho}}{\partial\vec{n}}$ is the regular density derivative in direction $\vec{n}=\vec{\nabla}H/\left|\vec{\nabla}H\right|$ of ocean depth H growth, q is the counter-gradient flux in this direction. This flux operates when the regular density gradient in the direction of the slope is zero. When the counter-gradient flux is $q>0$, it means that there is some downslope flux even for small opposite density gradient, which explains its name.

In our work [1], we used the results of regional modeling based on the SibPOM model of the Kara Sea in the Arctic [5], which resolves mesoscale eddies. We applied a number of well-known statistical methods to resulting dataset for analyzing the sensitivity of eddy fluxes to the characteristics of large-scale motion. Based on this approach, an approximating functional dependence of the eddy transport characteristics on the large-scale thermodynamic characteristics of the ocean was obtained. The resulting expressions for the parametrization dependencies were used in the framework of large-scale modeling of processes in the Arctic and the North Atlantic using the Siberian Coupled Ice-Ocean Model (SibCIOM) [6]. Thus, it was possible to assess the sensitivity of the model in relation to taking into account parametrized processes and to identify the main trends in the state of the ocean associated with the inclusion of cross-isobathic eddy transport in the Arctic region.

During the implementation of this methodology, a number of issues arose, some of which we will consider in this article.

2 Methodology

SibPOM sigma coordinate shelf model, which is a modification of the Princeton Ocean Model (POM) [7], includes the parameterization of vertical turbulent processes and the correction of the horizontal pressure gradient [8]. Its horizontal resolution is 3–4 km, which allows reproducing large mesoscale eddies. The results considered refer to a numerical experiment covering the period from September 2006 to September 2008 [5]. The time-averaged values of horizontal velocity components $\bar{u}$ and $\bar{v}$ along with the value of the density $\bar{\rho}$ at each point of the grid area, as well as the time-averaged values of the flux components $\overline{\rho u}$ and $\overline{\rho v}$, were obtained for each 10-days period. The values K and q were found using the least squares method as linear regression coefficients for all grid nodes in 50 km vicinity of any point (x_0, y_0, z_0) of five seas in Russian sector of Arctic including Barents, Kara, Laptev, East-Siberian and Chukchi Seas

$$K\langle\frac{\partial\bar{\rho}}{\partial\vec{n}}\rangle + q = -\langle\overline{\rho'\vec{u'}}\cdot\vec{n}\rangle. \tag{2}$$

Here $\langle\cdot\rangle$ denotes 50 km vicinity spatial averaging. Large-scale representation of high-resolution characteristics was found as linearization coefficients in the same vicinity in the form

$$\phi\left(x', y'\right) \approx Ax' + By' + C = \langle\frac{\partial\phi\left(x_0, y_0\right)}{\partial x}\rangle x' + \langle\frac{\partial\phi\left(x_0, y_0\right)}{\partial y}\rangle y' + \langle\phi\left(x_0, y_0\right)\rangle, \tag{3}$$

where ϕ is any of these characteristics, $(x', y') = (x - x_0, y - y_0)$. The coefficients A, B, and C were also found by the least squares method.

First of all, to develop earlier approach, this procedure should be also undertaken with respect to the vector $\vec{n} = (n_x, n_y)$, since the local bottom slope can have many mesoscale details that cannot be described in a large-scale model. Therefore, the direction of the slope in the equation (2) is defined as a large-scale characteristic and each component of this vector n_x and n_y should be calculated using linearization (3). This is one of the issue that should be raised in relation to [1].

Next, we will consider a certain set of hydrodynamic flow characteristics $(x_1, x_2, \ldots, x_N)$. As a result of linearization procedure (3), a sample was created, where each set of N large-scale hydrodynamic characteristics of the ocean was assigned the obtained values of K and q at each point of model domains. Alternatively, we could also consider direct dependence of downslope eddy mass flux

$$E = -\langle\overline{\rho'\vec{u'}}\cdot\vec{n}\rangle \tag{4}$$

on this characteristics, without parameterization like (1). Thus we have

$$(x_{1,i}, x_{2,i}, \ldots, x_{N,i}) \to (K_i, q_i, E_i), \quad i = 1, 2, \ldots, M \tag{5}$$

with sample volume M of about 700 million records.

The previously conducted cluster analysis revealed three most pronounced clusters among this set. Two of them, representing eddy formations in the area of influence of river runoff waters (9% of the sample) and in the region of a steep slope at the shelf boundary (3% of the sample), were not considered and now we can dwell on them in more detail. For the largest cluster (88% of the sample), it was found that among a number of large-scale characteristics, the values of K and q turned out to be the most sensitive to the value of the vertical component of the density gradient $\frac{\partial \rho}{\partial z}$, the component of the bottom density gradient in the direction of the bottom slope $\left.\frac{\partial \rho}{\partial \vec{n}}\right|_H$, and the value of the bottom slope itself $\left|\vec{\nabla} H\right|$.

2.1. Coefficient of Determination. The tightness of the linear dependence in the expression (2) is not the same when determining the coefficients K and q. Naturally, in the case when the linear dependence is close, the values of K and q are more reliable in their determination. Therefore, along with the coefficients themselves, it is appropriate to use a measure of the tightness of linear dependence in the form of a coefficient of determination

$$R = cor^2(-\overline{\rho' \vec{u'}} \cdot \vec{n}, \frac{\partial \bar{\rho}}{\partial \vec{n}}), \tag{6}$$

where *cor* stands for correlation coefficient of two arguments in 50 km vicinity. Thus, R is the square of the correlation coefficient of the values $-\overline{\rho' \vec{u'}} \cdot \vec{n}$ and $\frac{\partial \bar{\rho}}{\partial \vec{n}}$ in the vicinity of each sampling point. Statistically, the value of R expresses the part of the $-\overline{\rho' \vec{u'}} \cdot \vec{n}$ variability explained by its linear dependence on the value of $\frac{\partial \bar{\rho}}{\partial \vec{n}}$. We also excluded from consideration cases of a negative value of the diffusion coefficient K obtained as a result of applying (2) by setting the corresponding R value equal to zero. It is appropriate to take into account the specified determination coefficient at all subsequent stages, namely: when determining the degree of dependence on each other of large-scale flow characteristics, during clustering, sensitivity analysis, and when constructing the resulting functional dependencies of K and q on flow characteristics. It was not taken into account in our previous study [1], when all elements of the sample were treated equally.

When considering direct dependence of downslope eddy mass flux E on large-scale characteristics there is no need to relate it to defined determination coefficient. In this case we could consider it to be equal to a unit value for each sample element.

2.2. Large-scale flow characteristics. Table 1 presents a list of large-scale flow characteristics used in our analysis. Compared to [1] we extend it to include an angle α between density isosurface and sloping bottom, Rossby external radius of deformation $R_{\text{ext}} = \frac{\sqrt{gH}}{f}$, where g is the acceleration due to gravity, H is the ocean depth, and f is the Coriolis parameter, and squared

Rossby internal radius of deformation

$$R_{\text{int}}^2 = \frac{-\frac{g}{\rho}\frac{\partial \rho}{\partial z}H^2}{f^2}.$$

We use it squared to include possible unstable stratification of developing convection.

To find the least dependent characteristics in this list, we used factor analysis based on approach described in [10] to find the magnitude of the factor loadings. In this case, no assumptions are made about the nature of the variable distributions. To do this, first the sample of any variable x was rearranged in ascending order to form $x_j (j = 1, \ldots, M)$ and corresponding determination coefficient R_j sequences. Let $R_{\text{tot}} = \sum R_j$ (here and after if sum parameters are not indicated it is assumed summation for all $j = 1, \ldots, M$). For the variable x, we can determine its sample average $\bar{x}$ and variance $D(x)$, taking into account the coefficient of determination as follows

$$\bar{x} = \frac{\sum x_j R_j}{R_{\text{tot}}}, \tag{7}$$

$$D(x) = \frac{\sum R_j \left(x_j - \bar{x}\right)^2}{R_{\text{tot}}}. \tag{8}$$

In these expressions, the factor $\frac{R_j}{R_{\text{tot}}}$ has the meaning of the probability that the value x will be equal to x_j, taken to be proportional to the determination coefficient. Then the expressions will have the meaning of the expectation and variance of the x value.

The range of x_j change was divided into $L = 100$ intervals with about equal sum of R_j for each interval. That is we define boundaries $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{L+1}$ dividing the whole x range into intervals $I_i = [\tilde{x}_i, \tilde{x}_{i+1}]$ as follows

$$\begin{aligned}
&\tilde{x}_1 = \min x, \\
&\tilde{x}_{i+1} : \tilde{R}_i = \sum_{j:\ \tilde{x}_i < x_j < \tilde{x}_{i+1}} R_j \approx \frac{1}{L} R_{\text{tot}}, \\
&\tilde{x}_{L+1} = \max x.
\end{aligned}$$

For each of the intervals I_i, we can also determine the interval mean and inter-interval variance

$$\bar{x}_i = \frac{\sum\limits_{j \in I_i} x_j R_j}{\tilde{R}_i}, \tag{9}$$

$$\tilde{D}_x(x) = \frac{\sum\limits_{i=1}^{L} \tilde{R}_i \left(\bar{x}_j - \bar{x}\right)^2}{R_{\text{tot}}}. \tag{10}$$

The value $\tilde{D}_x(x)$ will tend to $D(x)$ with an increase in the number of intervals and will exactly match it if there is one element from the sample left in each interval.

If we take some other characteristic y and also find its mean, variance, interval mean, and inter-interval variance for same intervals I_i built for x, we have

$$\bar{y} = \frac{\sum y_j R_j}{R_{\text{tot}}}, \tag{11}$$

$$D(y) = \frac{\sum R_j \left(y_j - \bar{y}\right)^2}{R_{\text{tot}}}, \tag{12}$$

$$\bar{y}_i = \frac{\sum\limits_{j \in I_i} y_j R_j}{\tilde{R}_i}, \tag{13}$$

$$\tilde{D}_x(y) = \frac{\sum\limits_{i=1}^{L} \tilde{R}_i \left(\bar{y}_j - \bar{y}\right)^2}{R_{\text{tot}}}, \tag{14}$$

where x-subscript denotes that inter-interval variance is obtained using x intervals. Then the degree of dependence of y on x can be determined as a correlation ratio using the expression

$$\mathfrak{R}(y|x) = \frac{\tilde{D}_x(y)/D(y)}{\tilde{D}_y(y)/D(y)} = \frac{\tilde{D}_x(y)}{\tilde{D}_y(y)}, \tag{15}$$

where $\tilde{D}_y(y)$ is calculated the same way as $\tilde{D}_x(x)$ in (10) but for ascending set of y elements in sample. It means that $\mathfrak{R}(y|x)$ is a ratio of y inter-interval variance based on x intervals to y inter-interval variance based on y intervals.

Thus, considering the entire set of N state variables from the table 1, we can build a square matrix with elements

$$\mathfrak{R}_{ij} = \mathfrak{R}(x_i|x_j). \tag{16}$$

The diagonal elements of this matrix are equal to one, the elements of the row (first index) show the extent to which the value of x_i depends on other variables, and the elements of the column (second index) show the extent to which other variables depend on the value of x_j. What we need is to find a set of variables most independent, i.e. the corresponding row values to be small. On the other hand we are interested in variable to be more explaining, i.e. values in corresponding column to be greater. Let us define Row_i and Col_j as

$$\text{Row}_i = \left(\sum_{j=1}^{N} \mathfrak{R}_{ij}\right), \tag{17}$$

$$\text{Col}_j = \left(\sum_{i=1}^{N} \mathfrak{R}_{ij}\right). \tag{18}$$

Thus our desire to have variable most independed and more explaining could be expressed in terms of significance index Sig_i in form

$$\mathrm{Sig}_i = \frac{\mathrm{Col}_i}{\mathrm{Row}_i}. \tag{19}$$

Table 1 presents significance index for all selected characteristics obtained using SibPOM results [9] for five Arctic seas: Barents Sea (BS), Kara Sea (KS), Laptev Sea (LS), East Siberian Sea (ES), and Chukchi Sea (CS). As a final indicator we can use maximum possible significance index among these five seas. According to Table 1 the most valuable characteristics by this indicator are (in decending order) α, $\frac{\partial\rho}{\partial\vec{m}}$, R^2_{int}, $\left.\frac{\partial\rho}{\partial\vec{m}}\right|_H$, $\frac{\partial U}{\partial\vec{m}}$, $\frac{\partial V}{\partial\vec{m}}$, U, H, $\frac{\partial U}{\partial\vec{n}}$, $\frac{\partial V}{\partial\vec{n}}$, and R_{ext}.

However, some of these variables appear to be closely interrelated. For example, if we establish a chain of interdependencies, starting from the most significant value of α and taking into account only the values of $\mathfrak{R}_{ij}$ greater than 0.25, then for the Barents Sea we will get the system of relationships between state variables shown in Figure 1. In this figure the numbers next to the arrows connecting the elements of the system indicate in percentage how much the variance of one value is explained by the variability of the other. The number next to the variable designation indicates how much of the variance in that value is explained by the variability in the α value. Figure 1 shows that the variability of α is 88% related to the bottom slope $\left|\vec{\nabla}H\right|$, the variability of which is closely related to the density gradient in the direction of the slope. The value of R_{ext} is associated with the variability of α by only 37%. Nevertheless, it is very closely (99%) related to the value of H, on which the value of R^2_{int} largely (60%) depends. The structure of such relationships is somewhat different for different seas, however, its composition constantly contains the values $\left|\vec{\nabla}H\right|$, H, R_{ext}, $\left.\frac{\partial\rho}{\partial\vec{n}}\right|_H$, R^2_{int}, $\frac{\partial\rho}{\partial\vec{n}}$, $\frac{\partial\rho}{\partial z}$. Their dependence on the value of α averages 88, 52, 52, 39, 31, 25, and 22%, respectively. Three of them H, R_{ext} and R^2_{int} are also included in the previous list from which they can be excluded in this way. The quantities $\frac{\partial U}{\partial\vec{n}}$ and $\frac{\partial V}{\partial\vec{m}}$ are connected by a continuity equation, and the quantities $\frac{\partial U}{\partial\vec{m}}$ and $\frac{\partial V}{\partial\vec{n}}$ by the vorticity conservation law. Therefore, instead of these two pairs we chose velocity divergence $\mathbf{div}\,\vec{U} = \frac{\partial U}{\partial\vec{n}} + \frac{\partial V}{\partial\vec{m}}$ and its vorticity $\mathbf{curl}\,\vec{U} = \frac{\partial V}{\partial\vec{n}} - \frac{\partial U}{\partial\vec{m}}$. The value of $\left.\frac{\partial\rho}{\partial\vec{m}}\right|_H$ is important for shallow area or shelf break where it is closely related with $\frac{\partial\rho}{\partial\vec{m}}$. In deeper areas $\left.\frac{\partial\rho}{\partial\vec{m}}\right|_H$ is less important but $\frac{\partial\rho}{\partial\vec{m}}$ still is and therefore $\frac{\partial\rho}{\partial\vec{m}}$ is more relevant.

Despite the fact that, according to the Table 1, the significance of the horizontal component of the density gradient $\frac{\partial\rho}{\partial\vec{n}}$ is not large, we will still leave this characteristic among the state variables in case we will be considering the direct dependence of the eddy mass flux, so as not to exclude the possibility of identifying a dependence of the form (1).

TAB. 1. Large scale flow characteristics selected as possible state variables. Significance index is presented for Barentz Sea (BS), Kara Sea (KS), Laptev Sea (LS), East Siberian Sea (ES), and Chukchi Sea (CS). Combined index is a maximum index value.

Symbol	Description	Significance Index					
		BS	KS	LS	ESS	CS	Combined
H	ocean depth	1.2	1.1	0.9	1.0	1.0	1.2
σ	in situ depth relative to H	0.6	0.7	0.6	0.7	0.9	0.9
$\left\vert\vec{\nabla}H\right\vert$	bottom tilt	0.6	0.8	0.7	0.7	0.7	0.8
$\frac{\partial\rho}{\partial\vec{n}}$	in situ downslope density gradient component	0.7	0.8	0.9	0.7	0.8	0.9
$\frac{\partial\rho}{\partial\vec{m}}$	in situ along isobath density gradient component	1.2	1.2	1.7	1.1	1.4	1.7
$\frac{\partial\rho}{\partial\vec{n}}\Big\vert_H$	downslope density gradient component at the bottom	0.7	0.8	0.8	0.9	0.7	0.9
$\frac{\partial\rho}{\partial\vec{m}}\Big\vert_H$	along isobath density gradient component at the bottom	1.0	1.2	1.6	1.0	1.0	1.6
U	downslope horizontal velocity component	1.2	0.9	1.1	1.2	1.0	1.2
V	along isobath horizontal velocity component	1.1	0.9	0.8	0.8	0.9	1.1
$\frac{\partial U}{\partial\vec{n}}$	downslope velocity component of $\vec{\nabla}U$	1.1	0.6	0.9	1.1	0.9	1.1
$\frac{\partial U}{\partial\vec{m}}$	along isobath velocity component of $\vec{\nabla}U$	1.2	1.2	1.2	1.3	1.0	1.3
$\frac{\partial V}{\partial\vec{n}}$	downslope velocity component of $\vec{\nabla}V$	1.0	0.9	0.7	1.1	0.9	1.1
$\frac{\partial V}{\partial\vec{m}}$	along isobath velocity component of $\vec{\nabla}V$	1.1	0.8	1.1	1.2	0.9	1.2
$\frac{\partial\rho}{\partial z}$	vertical component of density gradient	0.8	0.7	1.0	1.0	0.9	1.0
α	angle between isopycnal and bottom surfaces in downslope direction	1.6	3.4	1.6	1.7	3.3	3.4
R_{ext}	external Rossby radius	1.1	1.0	0.9	0.9	0.9	1.1
R^2_{int}	squared internal Rossby radius	1.6	1.2	1.5	1.4	1.5	1.6

Finally, we have the following list of flow characteristics to be considered as variables:

$$(x_1, x_2, x_3, x_4, x_5, x_6) = \left(\alpha, \frac{\partial\rho}{\partial\vec{m}}, \mathbf{div}\,\vec{U}, \mathbf{curl}\,\vec{U}, U, \frac{\partial\rho}{\partial\vec{n}}\right) \tag{20}$$

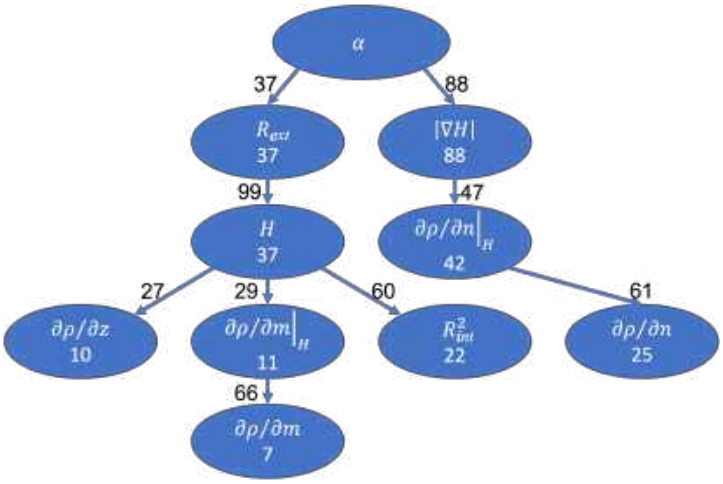

FIG. 1. A system of interdependent characteristics with a correlation ratio of at least 25% based on the results of calculations for the Barents Sea. The numbers next to the arrows connecting the elements of the system indicate in percentage how much the variance of following value is explained by the variability of the previous one. The number next to the variable designation indicates how much of the variance in that value is explained by the variability in the α value.

In our previous work, the α value was not considered, and based on the Fisher criterion, $\left|\vec{\nabla}H\right|$ and $\frac{\partial\rho}{\partial z}$ were among the most independent variables, which in our case is compensated by the angle α between the bottom slope and the isopycnal surface slope. We also used $\frac{\partial V}{\partial\vec{m}}$ and $\frac{\partial U}{\partial\vec{m}}$ instead of $\mathbf{div}\,\vec{U}$ and $\mathbf{curl}\,\vec{U}$.

For further cluster analysis, it is necessary to enter the distance between the sample elements in the assumed six-dimensional space. The value α, which is an angle in this case, is inconvenient to use, since the angles π and $-\pi$ are identical and, nevertheless, are separated by 2π. Therefore, we will consider α as an argument of the complex variable $z = \cos\alpha + i\sin\alpha$ and, for generality, we will assume that the correspondingly chosen variables are complex. That is, instead of $x_1 = \alpha$ in (20) we will use complex variable $x_1 = z$.

2.3. Clustering. The total sample, built on the results of a fine resolution simulation, contains elements consisting of a set (20) of selected parameters characterizing the large-scale motion and parameters describing the integral effect of mesoscale pulsations on the large-scale motion (K, q). In addition to this, we intend to use the coefficient of determination R and the value

of the resulting downslope eddy mass flux E. Since the nature of mesoscale movements can be completely different and refer to completely unrelated physical mechanisms, it makes sense to divide the entire sample into clusters, that is, into groups of the most closely related sample elements.

In this study, we use the so-called k-means method [11], in which belonging to a cluster is determined by the fact that the distance to its center is minimal among the centers of all clusters. The clustering procedure is iterative, after determining the belonging of elements to clusters, the center of each cluster is redefined in accordance with which elements are included in it. The iterations stop as soon as the composition of the clusters becomes unchanged.

An important issue in the implementation of the k-means method is the choice of the number of clusters k and the initial position of their centers. The choice of the number of clusters is based on the number of the sources of mesoscale motions, such as barotropic or baroclinic instabilities in the region of jet streams or near density fronts, as well as in regions of intense convective and wind mixing. We considered the values of k in a range from 2 to 25 as options searching for the resulting clusters to be more cohesive according to the relevant criteria, and also have a clear geographical localization, indicating a certain nature associated with this cluster physical processes (see below).

The values of the parameters characterizing the large-scale movement is a vector $(x_1, x_2, \ldots, x_N)$ in N-dimensional space, where N is equal to the number of these parameters (in our case, $N = 6$).

Since the parameters $(x_1, x_2, \ldots, x_N)$ are heterogeneous in nature, normalization is used to achieve their equivalence, that is, instead of a vector $\vec{x} = (x_1, x_2, \ldots, x_N)$, modified vector $\vec{X} = (X_1, X_2, \ldots, X_N)$ is used, where

$$X_i = \frac{x_i - \overline{x_i}}{\sigma_i}, \quad i = 1, \ldots, N \tag{21}$$

where mean value $\overline{x_i}$ and standard deviation σ_i are defined with the use of determination coefficient

$$\overline{x_i} = \frac{\sum_{j=1}^{\bar{M}} x_{i,j} \cdot R_j}{\sum_{j=1}^{\bar{M}} R_j}, \quad \sigma_i = \sqrt{\frac{\sum_{j=1}^{\bar{M}} |x_{i,j} - \overline{x_i}|^2 \cdot R_j}{\sum_{j=1}^{\bar{M}} R_j}}, \tag{22}$$

$\bar{M}$ is the total length of the sample, $x_{i,j}$ is the value of the i-th parameter in the j-th element of the sample and R_j corresponding determination coefficient. Note that $|\cdot|$ is a modulus of complex variable. For further analysis, the total number of sample elements is too big. The reduced sample that combines samples for all five seas under consideration was built on the basis of a random selection of elements with a probability proportional to the coefficient of determination. Following traditional approach, we should calculate the sum of all available coefficients over five seas

$$S = \sum_{k=1}^{5} \sum_{j=1}^{M_k} R_{k,j} = \sum_{i=1}^{M} R_i, \tag{23}$$

where index i corresponds to indexes k and j according to $i = j$ for $k = 1$ and $i = j + \sum_{l=1}^{k-1} M_l$ for $k > 1$. By use of commulative sum we could introduce

$$S_0 = 0, \quad S_m = \sum_{i=1}^{m} R_i,\ m > 1. \tag{24}$$

Then from the interval $[0, S]$ we can randomly select $\tilde{M}$ values of s_j ($j = 1, \ldots, \tilde{M}$) assuming a uniform distribution. Next, for each s_j we have to find corresponding m_j, so that

$$S_{m_j - 1} < s_j \leq S_{m_j}. \tag{25}$$

As a result of this procedure we would have $\tilde{M}$ elements of sample $\vec{X}_{m_j}$ ($j = 1, \ldots, \tilde{M}$). Unfortunately, for the chosen value $\tilde{M} = 4 \cdot 10^7$ and the sample volumes available for the seas, such a task requires too much CPU time. Therefore, we first found out how many elements $\tilde{M}_k$ of s_j fall on each of the seas. Then, for each k-th sea, the choice of these $\tilde{M}_k$ elements was randomly played. To do this, one of the elements l was randomly chosen and its membership in the sample was played with probability $R_{k,l}$. If the played event happen we consider $\vec{X}_{k,l}$ as a new member of reduced sample. And so on until the number of elements selected in this way to represent k-th sea reaches $\tilde{M}_k$. The implementation of this algorithm turned out to be significantly less expensive. The resulting sample was distributed over the seas BS, KS, LS, ES, and CS in the ratio (0.18, 0.16, 0.20, 0.24, 0.22), while the average coefficient of determination among the selected elements for these seas was the following values (0.154, 0.139 , 0.153, 0.156, 0.148).

Thus, we have obtained a combined sample consisting of $\tilde{M}$ elements, each j-th element of which consists of a set of parameters $x_{i,j}$, taken as variables, as well as values K_j and q_j from the representation (2), the corresponding coefficients of determination R_j, and the downslope eddy mass flux E_j values.

The clustering procedure is even more expensive in terms of CPU usage, therefore, for further actions, a group of $M = 2 \cdot 10^6$ elements is randomly selected from the selected $\tilde{M}$ elements, which is used to iteratively search for a cluster partition.

After normalization, the coordinates of the center of the l-th cluster are defined as

$$\overrightarrow{C_l} = (C_{1,l}, C_{2,l}, \ldots, C_{N,l}), \ \ C_{i,l} = \frac{\sum_{j \in S_l} X_{i,j} \cdot R_j}{\sum_{j \in S_l} R_j}, \tag{26}$$

where S_l is the set of sample elements belonging to the l-th cluster, M_l is the number of these elements, $X_{i,j}$ is the value of the normalized i-th parameter in the j-th sample element. After finding the centers of clusters, the sample element is considered as element of l-th cluster

$$\overrightarrow{X_j} = (X_{1,j}, X_{2,j}, \ldots, X_{N,j}) \in S_l \tag{27}$$

if a distance from it to l-th cluster center is minimal, i.e.

$$l: \; r_{j,l} = \min_{p=1}^{k} r_{j,p}, \;\; \text{where } r_{j,p} = \sqrt{\sum_{i=1}^{N} |X_{i,j} - C_{i,p}|^2}. \tag{28}$$

When split into two clusters, the initial position of cluster centers is determined by following the k-means++ [12] algorithm with regard of determination coefficient:

(1) The center of the first cluster is determined as a position of an element selected randomly with probability proportional to corresponding determination coefficient.
(2) The center of the second cluster is determined as a position of other element randomly selected with a probability proportional to the distance to the center of the first cluster multiplied by determination coefficient.

When splitting into $k > 2$ clusters, the centers of $k-1$ clusters are taken from the result of splitting into $k-1$ clusters, and the center of the additional k-th cluster is determined as a position of some element selected randomly with a probability proportional to the minimum among the distances to the known cluster centers multiplied by determination coefficient. The fact that the positions of the $k-1$ cluster centers are initially predistributed does not mean that their centers remain unchanged. Due to the iterative nature of the procedure, the introduction of an additional center leads to a redistribution of elements among clusters and to a change in the position of their centers. The resulting distribution is the one that, after 100 iterations, has the largest Dunn index [13].

The figure 2 shows the distribution of the best index values depending on the number of clusters, taking into account the coefficient of determination of a linear relationship (2). The best value of the index is in case of splitting into 2 and 3 clusters. However, this is the result of the selection of the most extreme clusters. In the first case (index value 4.12), there is a separation of a group of elements describing cases of unstable stratification in the upper layer of the ocean, resulting from cooling or salinity injection during ice freezing. In the second case (index value 0.73), processes are distinguished on a steep shelf slope, where the angle between the isopycnal surface and the bottom slope reaches extreme values. Therefore, it is worth considering partitions with a slightly larger number of clusters. In this case, the graph shows that the next two maxima fall on partitions with 6 and 9 clusters with index values of 0.67 and 0.54, respectively. The genealogy of the cluster origins is presented in Figure 3. It shows that the partition into 9 clusters contains clusters 5 to 9 inherited from clusters 2 to 6 from the partition into 6 clusters. Clusters 1 to 4 are based on the division of cluster 1 from a 6-cluster partition into four parts with some addition of elements from clusters 2 and 3. In the next section, we will consider what are the properties of the selected clusters.

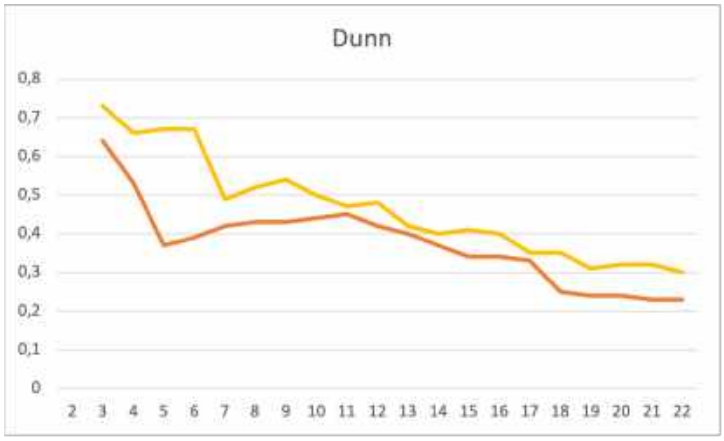

FIG. 2. Optimal Dunn index [13] for clusterization with different numbers of clusters: with account of determination coefficient (yellow line), and without it (red line).

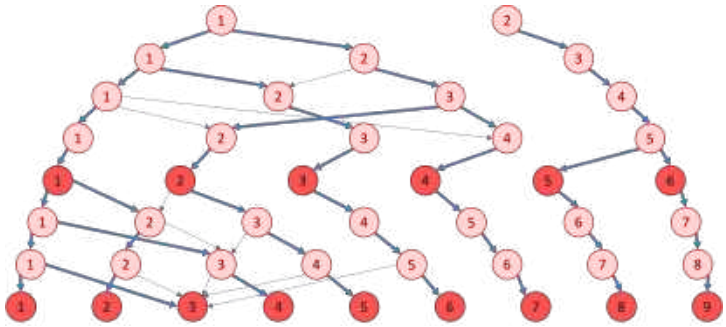

FIG. 3. Genealogical tree of inheritance of cluster properties when their number increases from 2 to 9. Bold arrows indicate which cluster from the previous partition the center of the subsequent partition cluster would belong to. Thin arrows indicate connections with other closest clusters, that is, with those whose distance to the center is no more than 2 times greater than the distance to the parent cluster. Clusters of 6 and 9-cluster partitions are highlighted in dark red.

The figure 2 also shows the distribution of the best index values depending on the number of clusters, provided that the presence of a deterministic linear relationship (2) is not decisive. That is, in all previous expressions, the value of the coefficients of determination was assumed to be identically equal to one. This makes sense in the case when we do not linearly relate the eddy mass transport with the value of the density gradient in the direction of the slope, but we are trying to find out the relationship between the value of this transport and the entire set of variables directly. As in the previous case, the

best index value is in the case of splitting into 2 and 3 clusters with index values of 5.91 and 0.64, which is also the result of selecting the same most extreme clusters. Considering partitions with a larger number of clusters, it can be noted that the second, somewhat smaller maximum of the index 0.45 takes place at the value of $k = 11$.

3 Clustering results

3.1. Six clusters accounting the determination coefficient. The Dunn index of this clustering is 0.67, that is, the minimum distance between cluster centers does not exceed two-thirds of the maximum standard deviations of cluster elements from their centers. Figures 4-8 show the distribution of some characteristics across clusters of this partition. Figure 4 demonstrates the geographic distribution of the vertically averaged coefficient of determination value and the average deviation of large-scale characteristics for each cluster relative to its overall mean in terms of the standard deviation of that characteristic. Figure 5 shows the distribution of eddy mass flux (4) for clusters, and figure 6 shows the distribution of kinetic energy of mesoscale eddies

$$EKE = \frac{1}{2}\rho_0 \left(\overline{u'^2} + \overline{v'^2}\right). \tag{29}$$

In addition, figures 7 and 8 demonstrate the eddy kinetic energy generation rate due to baroclinic

$$BC = -\frac{g^2}{\overline{N}^2 \rho_0}\left(\overline{u'\rho'}\frac{\partial\overline{\rho}}{\partial x} + \overline{v'\rho'}\frac{\partial\overline{\rho}}{\partial y}\right) \tag{30}$$

and barotropic

$$BT = -\rho_0\left(\overline{u'u'}\frac{\partial\overline{u}}{\partial x} + \overline{v'v'}\frac{\partial\overline{v}}{\partial y} + \overline{u'v'}\left(\frac{\partial\overline{v}}{\partial x} + \frac{\partial\overline{u}}{\partial y}\right)\right) \tag{31}$$

instabilities.

3.1.1. Cluster 1. This is the largest cluster, including 82.4% of the sample elements or 79.2% of determination coefficient sum. For this reason, its average values of large-scale characteristics are almost no different from the averages for the entire sample. Figure 4 demonstrates the ubiquitous presence of elements of this cluster among the seas with the highest values of determination coefficient near the shelf slope, as well as in the area of the New Siberian Islands. The elements are distributed evenly along the vertical, and the seasonal variability shows insignificant fluctuations of the total index (the sum of the determination coefficients of the cluster elements) within 15% with a maximum in April and a minimum in August.

The characteristics of the cluster differ slightly from the overall average values. The value of the angle α between the bottom and isopycnal surfaces has on average a small negative value, which indicates that the bottom slope at the points of the cluster elements is slightly greater than the slope of the isopycnal surfaces. Thus, this cluster is characterized by conditions that are

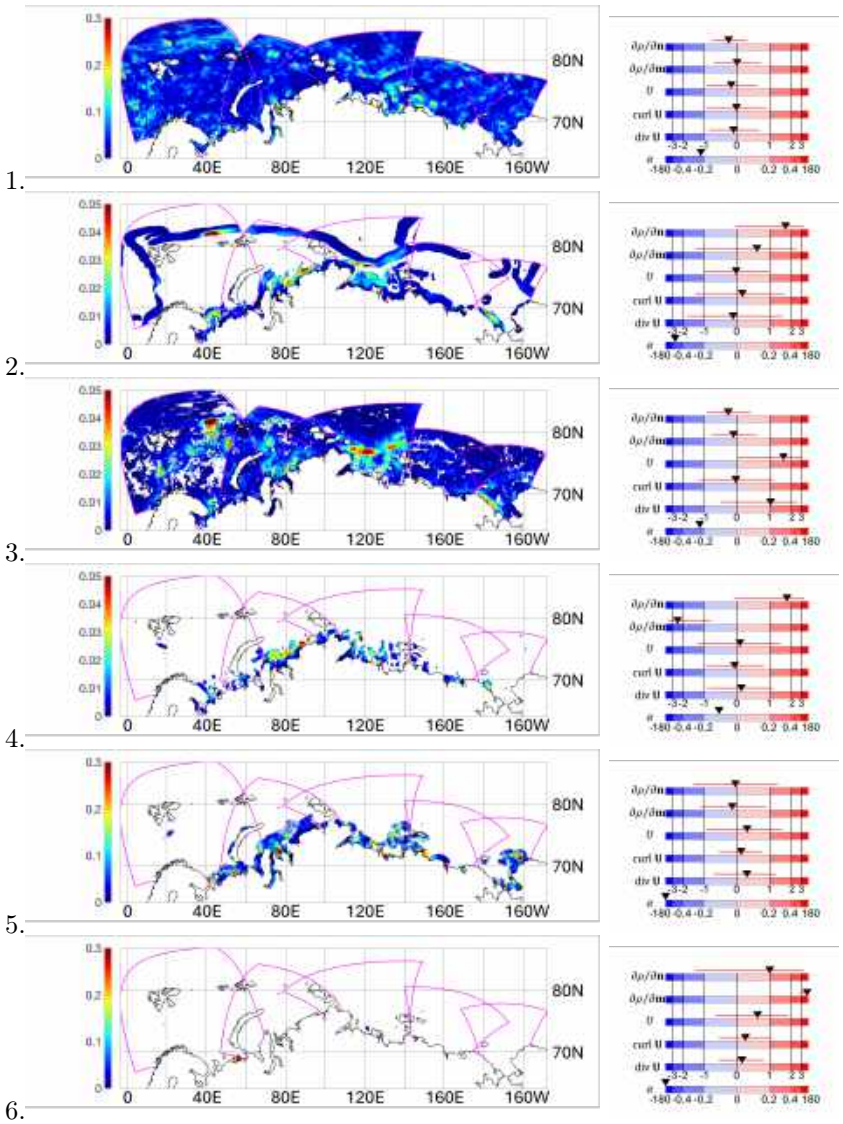

FIG. 4. Weighted average clustering with 6 clusters (cluster numbers are on the left of each panel): the geographic distribution of the vertically averaged coefficient of determination value (left column) and the average deviation of large-scale characteristics for each cluster relative to its overall mean in terms of the standard deviation (right column). The horizontal red lines at the top of each panel in the right column indicate the standard deviation range for each cluster. The last panel shows α in degrees.

not conducive to the development of cascading, the movement of dense water

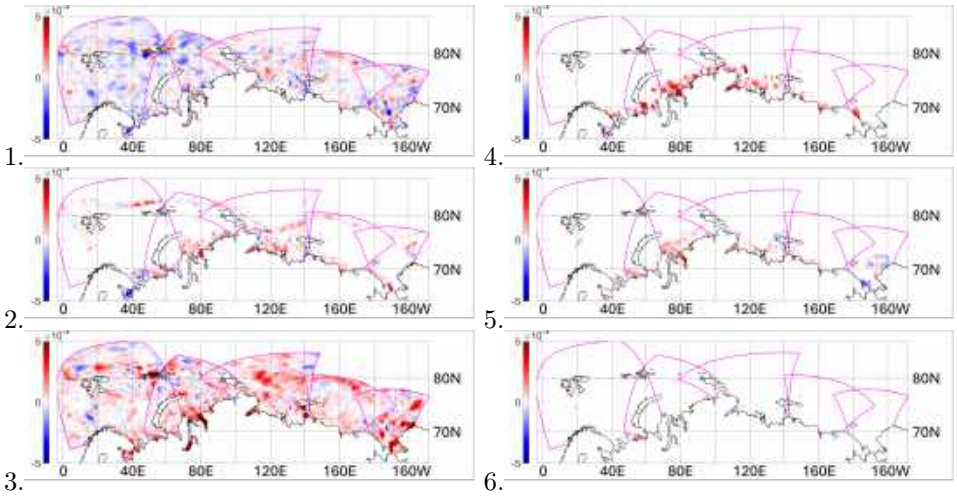

FIG. 5. Eddy mass flux (4) in kg/(m^2·s): the geographic distribution of the vertically averaged eddy mass flux for clustering with six clusters. Cluster numbers are on the left of each panel.

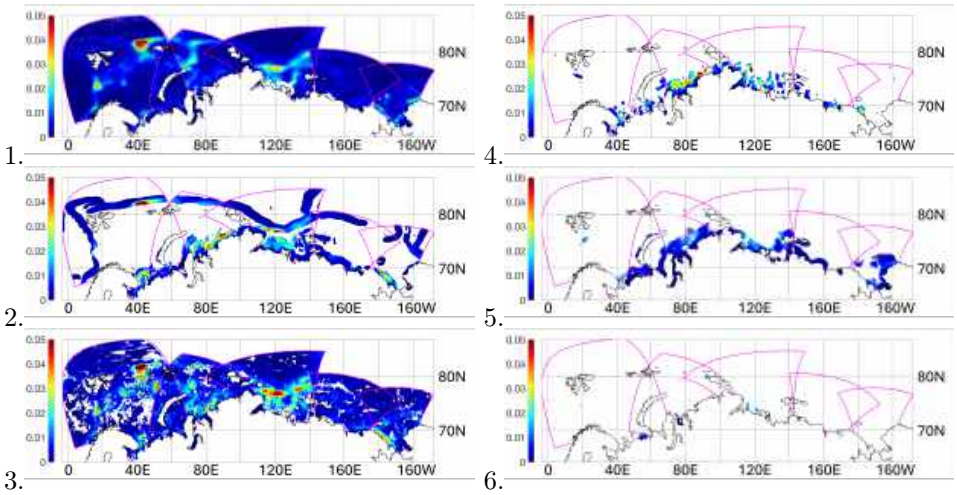

FIG. 6. Eddy kinetic energy (29) in J/m^3: the geographic distribution of the vertically averaged eddy kinetic energy for clustering with six clusters. Cluster numbers are on the left of each panel.

down along the inclined bottom. Moreover, the localization of this cluster coincides with the places of the highest values of the eddy mass flux (Fig. 5) with predominantly negative values in the Barents and Kara seas, which corresponds to the movement of mass up the slope. In the area of the Bering

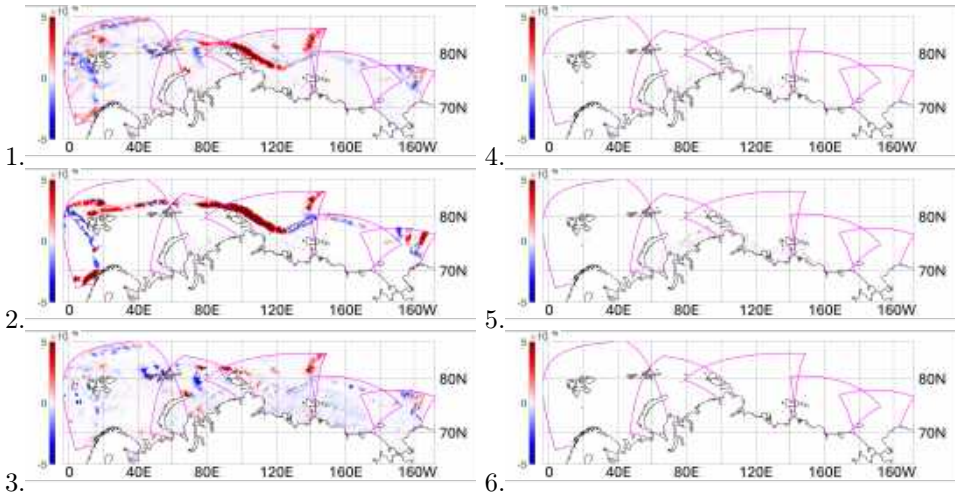

FIG. 7. Baroclinic generation rate of EKE (30) in W/m^3: the geographic distribution of the vertically averaged baroclinic generation rate of EKE for clustering with six clusters.

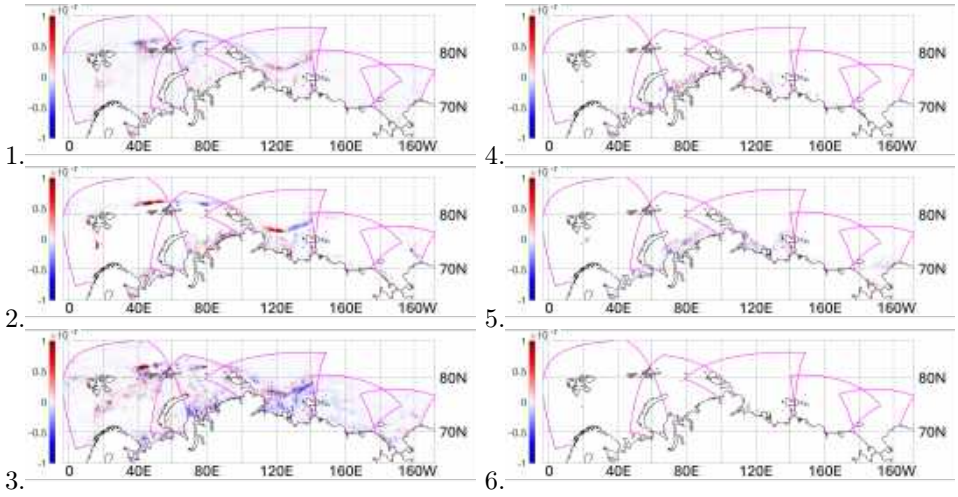

FIG. 8. Barotropic generation rate of EKE (31) in W/m^3: the geographic distribution of the vertically averaged barotropic generation rate of EKE for clustering with six clusters.

Strait in the Chukchi Sea, the elements of this cluster are dominated by those that correspond to intense and multidirectional eddy mass fluxes.

At the same time, the greatest eddy activity (Fig. 6) takes place along the trajectory of the Atlantic waters in the Arctic, namely along the coast of Scandinavia, along the axis of the underwater trenches - Barents Sea openning, the Franz Victoria, St. Anna and Voronin troughs , as well as on the shelf and shelf edge of the Laptev Sea and along the coasts where Pacific

waters from the Bering Strait spread. The eddies are mainly generated due to baroclinic instability in the area of the shelf slope between the mouth of the St. Anna Trough and the eastern margin of the Amundsen Basin near 125E in the Laptev Sea, as well as along the Lomonosov Ridge (Fig. 7). In the remaining sections of the topographic slope within this cluster, the eddy kinetic energy is mostly consumed. Barotropic instability (Fig. 8), as a source/sink of eddy kinetic energy, contributes 1-2 orders of magnitude less than baroclinic instability, and this is true for all detected clusters.

3.1.2. Cluster 2. This cluster includes 8.72% of sample elements or 10.1% of determination coefficient sum. The cluster elements are characterized by a more significant increase in density in the direction of the bottom slope $\frac{\partial \rho}{\partial \vec{n}}$ and a wider variability range of density in the perpendicular direction $\frac{\partial \rho}{\partial \vec{m}}$, of vorticity $\mathbf{curl}\,\vec{U}$ and flow divergence $\mathbf{div}\,\vec{U}$.

Geographically, the points of this cluster are located approximately in the 100-kilometer vicinity of the shelf slope, as well as along the coastline (Fig.4 and 5). In the first case, the most deterministic elements are located at depths of 500-2000 m. This explains the anomalously negative value of the α angle, since the slope of the isopycnals is not so great compared to the extreme bottom slope in these areas. In the case of a coastline, the largest number of cluster elements are located near the surface. The seasonal cycle for the cluster elements is more pronounced, since the total index varies within 50% of its average value with a maximum in August and a minimum in March.

The most intense eddies on the shelf slope (Fig. 6) are located northwest of Franz Josef Land near the mouth of the Franz Victoria Trough and northeast of Taimyr. In the coastal zone, eddy activity manifests itself at the exit from the White Sea, in the area of the western part of Taimyr and the Chukotka Peninsula. Baroclinic generation of eddy energy takes place almost along the entire shelf slope (Fig. 7) with the exception of the area on the approaches to the Lomonosov Ridge and in the Barents Sea openning area, where, on the contrary, eddy energy is consumed.

3.1.3. Cluster 3. The third largest cluster includes 7.34% of sample elements or 8.13% of determination coefficient sum. The cluster elements are characterized by an enlarged value of the U component of the current velocity in the direction of the slope, as well as an increased divergence of the horizontal velocity $\mathbf{div}\,\vec{U}$. Geographically, the points of this cluster are located throughout the entire water area, however, it can be noted that the most deterministic elements are noticeably more often located in the shelf zone (Fig.4). This apparently explains the increased value of the ageostrophic velocity component associated with Ekman pumping, the influence of which is much higher in the shelf zone. The most prominent areas are the vicinity of the Kola Peninsula, the mouth of the Pechora, the Gydan Peninsula, the vicinity of the Lena River delta and coastal areas to the east of it, as well as the vicinity of the Barrow Canyon north of Alaska.

The increased value of the ageostrophic component among the cluster elements explains the widespread predominance of the positive value of the eddy mass flux (Fig. 5) towards the topographic slope. The seasonal cycle for cluster elements is practically not manifested; the intra-annual variability in the total index of cluster elements does not exceed 10

The location of the most intense vortices (Fig. 6) as well as the areas of their generation (Fig. 7) repeats similar patterns for the first cluster, which is confirmed by their greater proximity according to the cluster inheritance scheme, shown in Fig. 3.

3.1.4. Clusters 4-6. Clusters 4 to 6 are a set of three small coastal clusters with volumes of 1.35, 0.2 and 0.01% of the sample or 2.17, 0.34 and 0.04% of determination coefficient sum. The largest of them, the 4th, is characterized by a large value of the density gradient components: both along the direction of the slope $\frac{\partial \rho}{\partial \vec{n}}$ and along isobath line $\frac{\partial \rho}{\partial \vec{m}}$ (Fig.4). Its seasonal variation demonstrates an almost twofold increase in the cluster volume in July compared to February. This is due to the amplification of density fronts in the coastal zone, caused by both seasonal heating and a summer increase in river flow.

Clusters 5 and 6 are distinguished by the fact that the value of the angle α between the slope of the isopycnals and the bottom slope is close to 180 degrees (Fig.4). It means the formation of unstable stratification. Therefore the seasonal cycle in both cases demonstrates the absence of elements of these clusters in the summer from May to September and their maximum number in the winter from October to March. At the same time, cluster 5 is characterized by a neutral value of other characteristics, and in cluster 6 increased components of the horizontal density gradient are distinguished (Fig.4).

In the case of cluster 4, the eddy mass flux down the slope is mainly positive and negative for clusters 5 and 6 (Fig. 5). The eddy kinetic energy (Fig. 6) is most significant at the boundary of these clusters, where its generation is also most active due to both baroclinic and barotropic instabilities (Fig. 7 and 8), and their contribution turns out to be the closest in order of magnitude.

3.2. Nine clusters accounting the determination coefficient. The Dunn index of this clustering is 0.54, which is considered a worse result compared to the previous clustering and means that the minimum distance between cluster centers does not exceed half of the maximum standard deviation of cluster elements from their centers. From the scheme of inheritance of cluster properties (Fig. 3) it follows that the first four clusters are mainly formed as a result of the division of the first cluster of the previous 6-cluster partition.

The parent's original cluster size was 82.39% of the sample size. The sizes of the four descendant clusters are 64.91, 9.76, 8.32 and 7.45%, respectively, and the first cluster differs practically little from its parent not only in volume, but also in other characteristics. The total size of these clusters

increased by 8.05% compared to the original cluster. This increase was mainly due to the reduction of neighboring clusters 2 and 3, the successors of which were clusters 5 and 6, the first of which decreased by 2.31% (instead of 8.72 it became 6.41%), and the second decreased almost four times (instead of 7.34 it became 1.75%) which also affected the change in its properties. Thus, the difference between a partition into 9 clusters and a partition into 6 clusters is that three new clusters 2-4 split off from the first cluster, while cluster 6 represents a significantly reduced cluster 3 of the previous partition. In the following sections, we will focus on what new properties the identified clusters 2-4 have and how the properties of cluster 6 have changed compared to its predecessor, cluster 3 of the previous 6-cluster partition.

Figure 9 shows the geographical location of the elements of clusters 2-4 and 6 in terms of the depth-averaged and time-averaged determination coefficient. It can be seen that, like the parent cluster, clusters 2-4 are located throughout the area, and their distribution does not depend greatly on depth (not shown). However, in contrast to cluster 1, the volumes of clusters 3 and 4 show significant seasonal variability, amounting to approximately a quarter of their value. In addition, if the direction of the vortex mass flow in cluster 1 of the previous partition was multidirectional, then in this partition cluster 1 demonstrates a predominantly positive direction (down the slope).

From the flux values presented in Figure 10 it is clear that cluster 3 also gives predominantly positive flux values, while cluster 4 gives mostly negative ones, and cluster 2 gives multidirectional fluxes. This fact of division into clusters with different behavior of mass fluxes justifies increasing the number of clusters from 6 to 9.

The distribution of kinetic energy of eddy motion for clusters 2-4, presented in Figure 11, is similar to that presented in Figure 6 for cluster 1, accounting some differences in the geographical location of the clusters themselves. A similar conclusion can be reached by comparing the distribution of baroclinic energy sources of these clusters.

A comparison of the properties of cluster 6 with its parent cluster 3 from the previous 6-cluster partition shows that the cluster has significantly reduced its volume and, as a result, the distribution of its elements has become more spotty (Fig. 9). In areas of the sea shelf remote from the coastline, islands and shelf slopes, cluster elements are almost completely absent.

The volume of the cluster began to significantly depend on the season, so that in the summer from May to October its volume is 2.5 times greater than the volume in the winter from December to March. The eddy mass flux has become less defined regarding its direction (Fig. 10) compared to the previous one. The distribution of kinetic energy, in principle, remained the same, taking into account the losses of a number of water areas, including the loss of eddies in the central part of the Laptev Sea, which moved mainly to cluster 3.

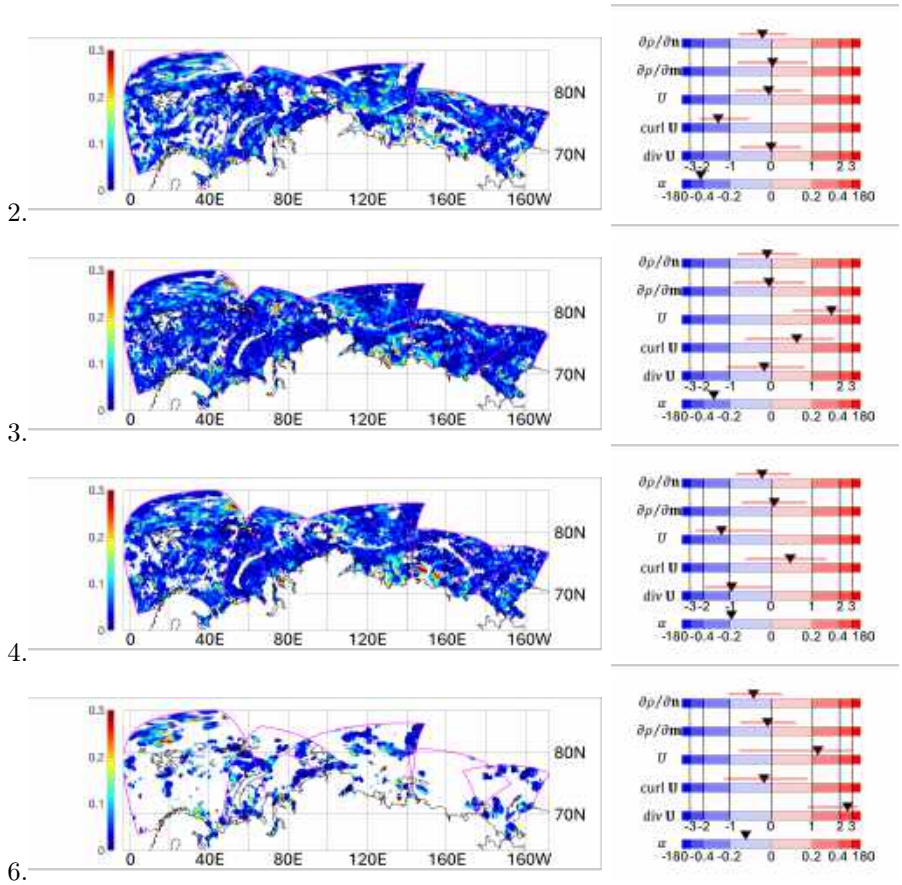

FIG. 9. Weighted average clustering with 9 clusters. Clusters 2-4 and 6 (numbers on the left): the geographic distribution of the vertically averaged coefficient of determination value (left column) and the average deviation of large-scale characteristics for each cluster relative to its overall mean in terms of the standard deviation of that characteristic (right column). The horizontal red lines at the top of each panel in the right column indicate the standard deviation range of the characteristics for each cluster. The last panel shows only the mean value of the angle α in degrees without reference to its overall mean and without the standard deviation.

Table 2 summarizes the clustering results, demonstrating the features of the large-scale characteristics of the cluster elements, the presence or absence of seasonal variation, and the presence of unidirectional eddy mass fluxes.

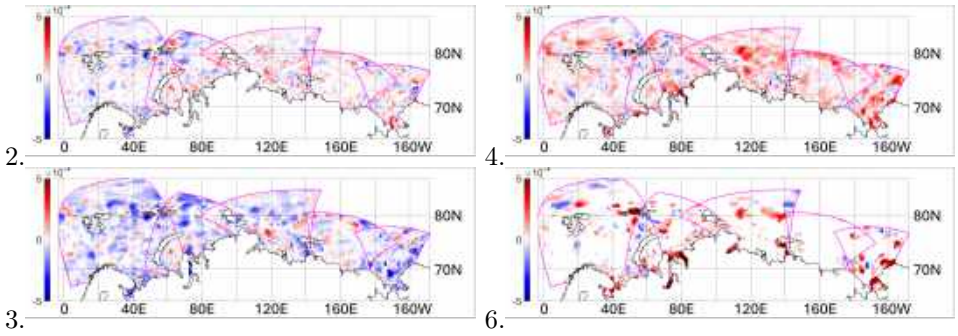

FIG. 10. 9 clusters partition. Eddy mass flux (4) of clusters 2-4 and 6 in kg/(m^2·s): the geographic distribution of the vertically averaged eddy mass flux for clustering with six clusters.

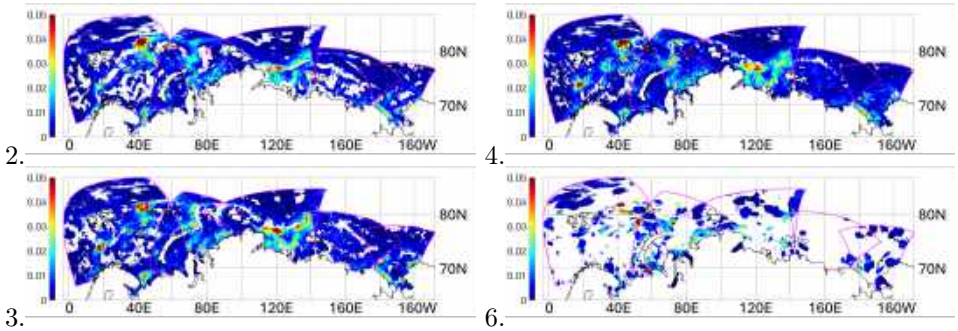

FIG. 11. 9 clusters partition. Eddy kinetic energy (29) of clusters 2-4 and 6 in J/m^3: the geographic distribution of the vertically averaged eddy kinetic energy for clustering with six clusters.

4 Conclusion

The presented work is in the context of the process of obtaining a parameterization of mesoscale mass transfer by eddy structures, described in the previous work [1]. To develop the approach, we consider here the results of numerical modeling using the SibPOM model of a wider range of Arctic seas, including (in addition to the Kara Sea [1]) the Barents, East Siberian, Chukchi and Laptev seas. A new algorithm for searching for the most independent large-scale flow characteristics is proposed, based on factor analysis of the sample, searching for correlation relationships between these characteristics, and calculating a composite index of the significance of the quantities under consideration. Since the idea of the approach is to search for coefficients of linear dependence in the form (2), factor analysis of sample elements was carried out taking into account the determination coefficient of this dependence.

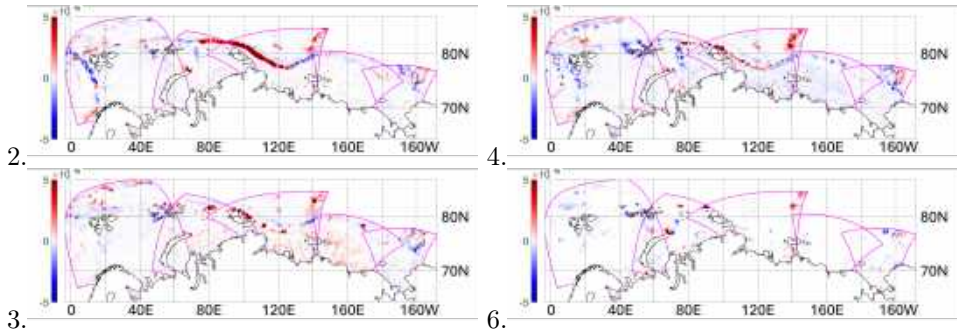

FIG. 12. 9 clusters partition. Baroclinic generation rate of EKE (30) of clusters 2-4 and 6 in W/m^3: the geographic distribution of the vertically averaged baroclinic generation rate of EKE for clustering with six clusters.

k		$\frac{\partial\rho}{\partial\vec{n}}$	$\frac{\partial\rho}{\partial\vec{m}}$	U	$\mathbf{curl}\,\vec{U}$	$\mathbf{div}\,\vec{U}$	α	Seasons	Dnsl EF
6	9								
1									−
	1								+
	2				−		−		
	3			+	+		−	w	+
	4			−		−		w	−
2	5	+					−	s	
3				+		+			+
	6			+		+		s	
4	7	+	−				+	s	+
5	8						!	w	
6	9	+	+				!	w	

TAB. 2. The results of clustering the selected parameters of large-scale movement for the number of clusters 6 and 9. "+"and "−"symbols mean that corresponding characteristics are significantly shifted in positive or negative direction. "Season"means seasonality of cluster: "w"or "s"mean that maximum cluster size is in winter or in summer. "Dnsl EF"means presence of downslope eddy mass flux if "+"and upslope if "−". "!"symbol means that α is an overturnung angle of ± 180.

The approach assumes that the sample obtained from numerical simulations with detailed resolution is divided into clusters. The purpose of such a division is to delimit physical processes of a mesoscale nature in the space of identified independent large-scale flow characteristics in order to subsequently search for acceptable parameterizations of the most significant of them. In

contrast to the approach outlined in the work [1], clustering used here is based on the determination coefficient of the linear dependence of the eddy mass flux in the direction of the topological slope on the density gradient component in this direction, since the idea of the approach is to search for the coefficients of this linear dependence in the form (2).

The paper presents the results of such clustering, showing that local maxima of the Dunn index are achieved when choosing the number of clusters equal to 6 or 9. It describes the main features of the elements of the resulting clusters, the nature of the eddy fluxes and their energy significance.

The work was carried out under project No. 123081400010-2 with the support of the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-03-2023-506/1). It uses the results of numerical experiments and their analysis obtained within the framework of the continued RSF project No. 20-11-20112.

References

[1] G. Platov, D. Iakshina, E. Golubeva *Parametrization of Eddy Mass Transport in the Arctic Seas Based on the Sensitivity Analysis of Large-Scale Flows*, Water, **15** (2023), 472. doi: 10.3390/w15030472

[2] J. W. Deardorff, G. E. Willis *Turbulence within a baroclinic laboratory mixed layer above a sloping surface*, J. Atmos. Sci, **44** (1987), 772–778.

[3] V. N. Lykossov *K-theory of atmospheric turbulent planetary boundary layer and the Boussinesq's generalized hypothesis*, Sov. J. Numer. Anal. Math. Modeling, **3** (1990), 221–240.

[4] V. N. Lykossov, G. A. Platov *A numerical model of interaction between atmospheric and oceanic boundary layers*, **7** (1992), 419–440. doi: 10.1515/rnam.1992.7.5.419

[5] G.A. Platov, E.N. Golubeva *Interaction of Dense Shelf Waters of the Barents and Kara Seas with the Eddy Structures*, Physical Oceanography, **26** (2019), 484–503. doi:10.22449/1573-160X-2019-6-484-503

[6] E.N. Golubeva, G.A. Platov *Numerical Modeling of the Arctic Ocean Ice System Response to Variations in the Atmospheric Circulation from 1948 to 2007*. Izvestiya, Atmospheric and Oceanic Physics, **45** (2009), 137–151. doi: 10.1134/S0001433809010095

[7] A.F. Blumberg, G.L. Mellor *A Description of a Three-Dimensional Coastal Ocean Circulation Model*. In: Three-Dimensional Coastal Ocean Models, N. S. Heaps (ed.), Coastal and Estuarine Sciences, **4**, American Geophysical Union, Washington, D.C., 1–16, 1987.

[8] G.A. Platov, J.F. Middleton *Notes on Pressure Gradient Correction*, Bulletin of the Novosibirsk Computing Center, **7** (2001), 43–58. Available at: https://nccbulletin.ru/files/article/platov_3.pdf [Accessed: 31.08.2022].

[9] G. Platov, E. Golubeva *Characteristics of mesoscale eddies of Arctic marginal seas: Results of numerical modeling*, IOP Conference Series: Earth and Environmental Science, **11** (2020), 012009. doi: 10.1088/1755-1315/611/1/012009

[10] G. Berkooz, P. Holmes, J. L. Lumley *The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows*, Annual Review Of Fluid Mechanics, **25** (1993), 539–575. doi: 10.1146/annurev.fl.25.010193.002543

[11] S. P. Lloyd *Least squares quantization in PCM*, IEEE Transactions on Information Theory, **28** (1982), 129–137. doi: 10.1109/TIT.1982.1056489

[12] D. Arthur, S. Vassilvitskii *k-means++: the advantages of careful seeding*, Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics Philadelphia, PA, USA. 1027–1035, 2007.
[13] J. Dunn *Well separated clusters and optimal fuzzy partitions*, Journal of Cybernetics, **4** (1974), 95–104. doi:10.1080/01969727408546059

Gennady Alexeyevich Platov
Institute of Computational Mathematics and Mathematical Geophysics SB RAS,
pr. Lavrentieva, 6,
630090, Novosibirsk, Russia
Email address: plat@ommfao.sscc.ru

Elena Nikolaevna Golubeva
Institute of Computational Mathematics and Mathematical Geophysics SB RAS,
pr. Lavrentieva, 6,
630090, Novosibirsk, Russia
Email address: elen@ommfao.sscc.ru