УДК 551.465

В. Н. Коморин, *к.г.н., с.н.с.,* **Л. В. Мацокин,** *к.т.н.,* **А.В. Кочеткова**
НИУ "Украинский научный центр экологии моря"

МЕТОД УСВОЕНИЯ АРХИВНОЙ ИНФОРМАЦИИ В МОДЕЛЯХ ДИНАМИКИ МОРЯ

На основе алгоритмов фильтрации Р. Калмана разработан метод усвоения архивной информации в математических гидродинамических моделях моря с предварительной обработкой данных с помощью алгоритма автоматической кластерзации.
Ключевые слова: *мониторинг морской среды, модели динамики моря, фильтрация Р. Калмана, усвоение архивной информации, автоматическая кластеризация.*

Введение. Значительный объем первичной информации, накопленной Украинским Научным Центром Экологии моря (УкрНЦЭМ) при проведении мониторинга морской среды, позволяет использовать результаты натурных наблюдений для более углубленного изучения состояния моря посредством усвоения архивных данных (результатов длительного периода наблюдений) детерминированными и стохастическими моделями динамики моря. Основным источником информации УкрНЦЭМ служат гидрологические полигонные съемки, выполняемые научно-исследовательскими судами (НИС) в различные временные периоды, а также математические гидродинамические модели. При картировании результатов мониторинга полученные данные относят к одному и тому же моменту времени, а затем интерполируют между соседними гидрологическими станциями. Однако по таким картам можно получить лишь качественное представление о пространственной изменчивости полей, поскольку они подвергаются значительному осреднению, что приводит к смещению изолиний полей относительно реального распределения характеристик водных масс. При этом, существенно искажаются статистические оценки полей и как следствие результаты вычислений в численных схемах решений систем уравнений движения [1].

Постановка задачи. Целью работы является разработка метода по усвоению результатов натурных наблюдений за состоянием моря и получение оптимальных оценок гидрологических и гидрохимических характеристик в узловых точках модельной сетки и в точках поля, где измерения не проводились [2].

В информационном массиве, представляющем результаты длительного периода наблюдений, данные за одну съемку, относительно основного массива данных, неэквидистантны во времени и в пространстве (из-за изменений координат гидрологических станций каждого экспедиционного рейса). Для получения более полной «усредненной» картины динамики океанографических полей, необходимой для прогноза их изменчивости, следует производить соответствующие прогностические расчеты и картирование полей по информации, полученной как за период единичной съемки, так и за все предшествующие периоды. Фактически, слежение за динамикой полей предполагает чередование этапов их прогноза, полученного посредством математического моделирования и адаптации его оценок к натурным наблюдениям (усвоение данных конкретного экспедиционного рейса) [2].

Входные данные и методы их обработки. Входными данными являются данные комплексного морского экологического мониторинга и результаты модельных расчетов. Коррекция расчетов математической модели данными измерений представляет собой одну из наиболее актуальных задач в современной математической геофизике. В мировой практике такие задачи получили название "задачи усвоения".

Одним из подходов решения задач усвоения является схема, основанная на статистической теории оценивания и фильтрации процессов на фоне "шума" с известными статистическими свойствами. Этот подход, получивший название "фильтрации Калмана" (Kalman filtering) достаточно просто реализуем и устойчив к малым возмущениям наблюдаемых данных [3].

Метод усвоения архивной информации в гидродинамических моделях. Отличительной особенностью алгоритмов фильтрации является условный характер осреднения уравнений изменчивости океанографических характеристик и наличие уравнений для вторых моментов распределений вероятностей ошибок прогнозов. Следует отметить, что использование фильтрации Р.Калмана правомерно лишь в случае линейной динамики, где идея условного осреднения уравнений моделей морской среды базируется на свойствах гауссовско-марковских случайных функций [2]. Свойство линейности эквивалентно допущению о нормальном (гауссовском) распределении вероятностей, а свойство марковости о зависимости всех последующих x_i значений процесса от x_{i-1} предшествующих значений. Гауссовско-марковские свойства модели позволяют представить прогностические оценки процесса в виде [4]

$$\hat{X}_k = \Phi_{k,k-1}\hat{X}_{k-1} + K_k\left[Z_k - \Phi_{k,k-1}\hat{X}_{k-1}\right], \qquad (1)$$

где $\Phi_{к,к-1}$ - коэффициент связи (переходная функция); $K_к$- весовой коэффициент, определяющий точность прогноза; $Z_к$- вектор измерений параметров морской среды.

Как видно из (1) уравнения оптимальной фильтрации представляет собой сумму детерминированной и случайной составляющих, где первое слагаемое - детерминированный прогноз процесса, а второе – стохастическая поправка к прогнозу. Качество прогноза характеризуется дисперсией ошибки прогноза

$$\sigma_k^2 = E\left\{\left(X_k - \hat{X}_k\right)^2\right\}, \qquad (2)$$

где: E - символ математического ожидания; $X_к$ - действительные значения параметров процесса.

Правильный выбор весового коэффициента в уравнении (1) обращает в минимум дисперсию, что означает максимально возможную точность предсказания. Нахождение же наилучшего значения весового коэффициента есть задача оптимальной фильтрации векторного процесса по его наблюдениям. Основной проблемой реализации алгоритма оптимальной фильтрации является большая размерность ковариационных матриц ошибок прогноза. Решение данной проблемы состоит либо в использовании упрощенных операторов для расчета матриц ковариаций либо применением ансамблевого подхода (принятие гипотезы эргодичности случайных полей ошибок прогноза). Подобные алгоритмы фильтрации получили название субоптимальных алгоритмов Р. Калмана, применение которых позволяет более осознанно подойти к моделированию процессов изменчивости параметров моря. Отдельно следует отметить, что одним из условий применимости динамико-стохастического подхода к решению задач усвоения данных является отсутствие тренда в характере изменчивости параметров морской среды.

Поскольку изменчивость параметров моря существенно зависит от глубины, а взаимосвязь между параметрами подвержена значительным колебаниям, для практической реализации работы алгоритма усвоения наиболее целесообразно использование субоптимальных алгоритмов фильтрации, производя расчеты отдельно для каждого горизонта наблюдений, совпадающего со слоем расчетной сетки. Выбор числа параметров для фильтрации следует приводить исходя из степени линейной взаимосвязи, определяемой коэффициентами корреляции между параметрами моря.

Перед осуществлением фильтрации Р.Калмана, для соблюдения допущений о

гауссово-марковских свойствах моделируемых процессов, необходима предварительная обработка архивной информации в совокупности с результатами модельных расчетов. Суть предварительной обработки состоит в выделении однородных зон в пределах, которых распределение значений параметров моря относится к нормальному закону распределения, при отсутствии трендовой составляющей в сформированных для фильтрации рядах.

Для оценки однородности наиболее приемлем метод автоматической кластеризации, позволяющий выделить однородные зоны (кластеры) в многомерном пространстве параметров моря, при не задании априорного числа кластеров. В УкрНЦЕМ разработан метод автоматической кластеризации, основанный на переходе из исходного многомерного пространства в пространство первых главных компонент корреляционной матрицы параметров моря и оценивании однородности объектов по критическому радиусу, численные значения которого определяются величиной стандартного отклонения объектов от центра тяжести *i*-того кластера. В качестве центра тяжести кластера выбирается объект с минимальным суммарным эвклидовым расстоянием до остальных объектов. Объекты с расстоянием меньшим или равным критическому, считаются однородными, им присваивается соответствующий *i*-тый номер кластера. В последующих расчетах объекты, попавшие в *i*-тый кластер, не участвуют. Критерием однородности служит критический радиус, величина которого ограничена односторонним доверительным интервалом для нулевого среднего (центра тяжести кластера)

$$R_i^{krit.} = \frac{t_\alpha \sigma_i}{\sqrt{n_i}}, \qquad (3)$$

где: t_α - критерий Стьюдента при заданном уровне значимости α; σ- стандартное отклонение радиусов объектов от центра тяжести кластера; n_i- количество объектов при расчете критического радиуса.

Результатом работы алгоритма является серия кластеров, плотность объектов в которых может служить косвенным критерием качества планирования сетей гидрологических станций в проведенных экспедициях. Поскольку в плотных кластерах изменчивость переменных незначительна, то можно считать, что в пределах границ таких кластеров количество измерений достаточно для оценки состояния моря по выбранным параметрам, в то время как в разряженных кластерах значительные градиенты параметров свидетельствуют о некоторой неопределенности характера их изменчивости, для устранения которой необходима дополнительная информация.

После завершения процедуры автоматической кластеризации следующий этап субоптимального алгоритма усвоения состоит в формировании рядов каждого показателя (пакета данных), включая координаты и значения показателей детерминированной модели. Причем первыми членами рядов служат значения параметров моря, полученные на гидрологических станциях наиболее удаленных от точки модельной сетки (как по географическим координатам, так и по времени). При достаточно большой длине ряда (более 20 значений) проводится проверка ряда на наличие линейного тренда.

Процедуру фильтрации наиболее рационально проводить отдельно для каждого параметра, характеристикой погрешности измерений которого является дисперсия. Обобщением дисперсии для совокупности скалярных величин служит ковариационная матрица, диагональные элементы которой являются дисперсии составляющих вектора, а недиагональные – ковариации, характеризующие взаимосвязь между парой составляющих.

На следующем шаге алгоритма усвоения (для отдельных параметров) определяется величина весового коэффициента при вариабельности измерений

параметра, определяемой величиной дисперсии (для соблюдения размерности параметра принято стандартное отклонение) в *i*-том кластере. Алгоритм последовательно обрабатывает скалярные величины, начиная с гидрологической станции, наиболее близко расположенной к точке модельной сетки *i*-того кластера. На начальном шаге происходит вычисление веса поправки - невязки между модельным значением параметра и результатом наблюдения на близлежащей станции, с последующим уточнением значений параметра и его дисперсии. Вес поправки на первом шаге *i*-того кластера определяется в виде

$$K_0^i = \frac{\sigma_0^i}{\left(\sigma_0^i + \sigma_1^i\right)}, \tag{4}$$

где: σ_0^i - стандартное отклонение модельной оценки параметра; σ_1^i - стандартное отклонение параметра в *i*-том кластере модельной точки и близлежащей к ней гидрологической станции.

Исправленное модельное значение параметра составит

$$\left(C_{м.испр.}^i\right)_1 = \left(1 - K_0^i\right)C_м^i + K_0^i C_1^i, \tag{5}$$

где: $C_{м.испр.}^i$ - исправленное модельное значение параметра в *i*-том кластере; $C_м^i$ - начальное модельное значение параметра; C_1^i - значение параметра на близлежащей к модельной точке станции.

Неопределенность следующей исправленной модельной оценки будет определена в виде

$$\left(\sigma_1^i\right)_{испр.} = \left(1 - K_0^i\right)\sigma_1^i. \tag{6}$$

Следующее значение весового коэффициента соответственно составит

$$K_1^i = \frac{\left(\sigma_0^i\right)_{испр.}}{\left(\left(\sigma_1^i\right)_{испр.} + \sigma_1^i\right)}, \tag{7}$$

где: σ_1^i - стандартное отклонение ряда, включающего модельное значение, а также значения параметров на близлежащей к модельной точке и предшествующей близлежащей станции. Исправленное модельное значение параметра будет определено в виде

$$\left(C_{м.испр.}^i\right)_2 = \left(1 - K_1^i\right)\left(C_{м.испр.}^i\right)_1 + K_1^i C_2^i, \tag{8}$$

где: C_2^i - значение параметра на гидрологической станции, предшествующей близлежащей станции.

Неопределенность следующей исправленной модельной оценки $\left(C_{м.испр.}^i\right)_2$ составит

$$\left(\sigma_2^i\right)_{испр.} = \left(1 - K_1^i\right)\sigma_2^i. \tag{9}$$

Аналогично зависимостям (4 – 9) проводятся дальнейшие расчеты весового коэффициента и окончательного исправленного модельного значения.

В качестве примера работы алгоритма в таблице приведены результаты усвоения данных для одного из кластеров для температуры, солености и растворенного кислорода весеннего периода поверхностного слоя северо-западной части Черного

моря [5], в совокупности с результатами модельных расчетов, полученных с помощью модели POM (Princeton Ocean Model) [6].

Таблица - Результаты работы алгоритма усвоения данных (данные натурных наблюдений расположены по мере удаления по времени и координатам от модельной точки)

Шаг алгоритма	Параметр	Длина ряда	Ст.откл. рядов	Весовой коэфф.	Исправленное модельное значение параметра	Исправл. ст. откл.
1	*2*	*3*	*4*	*5*	*6*	*7*
Температура, °С						
	15,93(модель	-	3,860			
1	12,70	2	3,698	0,511	15,259	1,889
2	12,75	3	3,005	0,386	14,291	1,159
3	12,64	4	2,617	0,307	13,783	0,804
4	14,92	5	2,303	0,259	14,077	0,596
5	12,83	6	2,133	0,218	13,805	0,466
Соленость, е.п.с.						
	11,91(модель	-	2,990			
1	11,079	2	0,587	0,836	11,215	0,490
2	11,708	3	0,433	0,531	11,511	0,229
3	12,182	4	0,469	0,328	11,731	0,154
4	11,581	5	0,411	0,273	11,690	0,112
5	12,465	6	0,485	0,188	11,835	0,091
Кислород, мл\л						
	7,87(модель)	-	1,36			
1	6,69	2	0,834	0,619	7,139	0,518
2	6,90	3	0,629	0,452	7,031	0,284
3	6,92	4	0,527	0,345	6,992	0,186
4	6,57	5	0,513	0,266	6,880	0,137
5	7,09	6	0,461	0,229	6,928	0,106

Как следует из таблицы величина весового коэффициента уменьшается по мере удаления от координат модельной точки. Окончательным модельным значением после усвоения архивных данных следует считать: температуру – 13,805 °С, соленость – 11,835 е.п.с., кислород – 6,928 мл\л.

Подобные расчеты проводятся для всех точек модельной сетки в каждом из выделенных кластеров, с последующей оценкой качества прогноза (исправленных модельных значений) для всей расчетной сетки по критерию качества (2). В результате, построенные изолинии полей по исправленным модельным значениям будут отражать не только детерминированную составляющую, полученную с помощью математического моделирования, но и стохастическую, оцененную посредством усвоения моделью архивного материала. При усвоении многомерной информации последовательность расчетов аналогична зависимостям (4-9), где вместо дисперсий используются ковариационные матрицы центрированных и нормированных данных.

Выводи. Крайне важным вопросом правомерности использования предлагаемого алгоритма является динамическая неустойчивость водных масс [1]. Однако, как показали

результаты натурных исследований, термохалинный и многомерный статистический анализ значительного объема информации [5], в характере распределения водных масс, в частности северо-западной части Черного моря, в не зависимости от года, в каждый гидрологический сезон образуются устойчивые однородные зоны (районы моря), в пределах которых (в многолетний период) можно абстрагироваться от динамической неустойчивости вод и выделить кластеры, в которых распределение исследуемых характеристик близко к нормальному. Это позволяет использовать фильтрацию данных в не зависимости от времени проведенных исследований. Таким образом, можно получить обобщенные за многие годы поля, используемые для оценки степени их изменчивости в пределах каждого экспедиционного рейса, что позволяет сигнализировать, в случаях значительных изменений границ полей, о доминирующем влиянии позитивного или негативного факторов на изменчивость структуры водных масс. Правомерность же использования предлагаемого алгоритма усвоения основана одновременно на двух концепциях однородности многомерных данных – вероятностно-статистической и структурной (геометрической) [7]. Суть первой состоит в разделении исходной совокупности данных на однородные группы, каждая из которых представляет собой смесь из нескольких нормально распределенных выборок случайных величин, что позволяет внутри каждой группы использовать алгоритмы фильтрации Р. Калмана. Суть структурной концепции состоит в выделении компактных групп объектов, удаленных друг от друга на некотором, заранее заданном расстоянии.

Задачи дальнейшего исследования: На основании предложенного метода планируется провести анализ данных, полученных Украинским Научным Центром Экологии Моря в рамках проведения государственного комплексного мониторинга за период с 1990 по 2009 гг.

Список литературы

1. *Бенземан В.Ю.* Пространственно-временная изменчивость гидрофизических полей океана. - СП.:- 2004. - 210 с.
2. *Тимченко И.Е., Ярин В.Д., Васечкина Е.Ф., Игумнова Е.М.* Системный анализ морской среды / НАН Украины // Морской гидрофизический институт - С.:- 1996. - 224 с.
3. *Калман Р., Фалб П., Арбиб М.* Очерки по математической теории систем. - М.: Мир,1971. - 400с.
4. *Еремеев В.Н., Игумнова Е.М., Тимченко И.Е.* Моделирование эколого-экономических систем. - Севастополь.: 2004. - 322 с.
5. *Орлова І. Г., Павленко М. Ю., Український В. В., та ін.* Гідрологічні та гідрохімічні показники стану північно-західного шельфу Чорного моря / Під редакцією І. Д. Лоєва. - К.: КНТ, 2008. – 616с.
6. *Коморін В. М., Попов Ю. І., Український В. В.* Оцінка мінливості гідродинамічних характеристик північно-західного шельфу Чорного моря/Вісник Одеського державного екологічного університету. – 2008. – Випуск 5. – с. 188 - 201
7. *Мандель И.Д.* Кластерный анализ. - М.: Финансы и статистика, 1988. - 176с.

Метод засвоєння архівної інформації в моделях динаміки моря.

В. М. Коморін, Л. В. Мацокін, А.В. Кочеткова.

На основі алгоритмів фільтрації Р. Калмана розроблено метод засвоєння архівної інформації в математичних гідродинамічних моделях моря з попередньою обробкою даних за допомогою алгоритму автоматичної кластеризації.

Ключові слова: *моніторинг морського середовища, моделі динаміки моря, фільтрація Р. Калмана, засвоєння архівної інформації, автоматична кластерізація.*

Methods of mastering of the archival information in dynamic models of the sea.

V.N. Komorin, L.V. Matsokin, A.V. Kochetkova.

On the basis of algorithms filtration R. Kalmana the method of mastering of the archival information in mathematical hydrodynamic models of the sea was developed with preliminary data processing with the help of algorithm automatic clusterization.

Keywords: *monitoring of the sea environment, dynamic models of the sea, R. Kalman's filtration, mastering of the archival information, automatic clusterization.*