

*Е.П. Истомин, В.В. Новиков, О.Н. Колбина, А.Ю. Сидоренко, С.Ю. Степанов*

## **СЛОЖНАЯ ИНФОРМАЦИОННАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ РИСКОВ С ПРИМЕНЕНИЕМ ФИЛЬТРА КАЛМАНА–БЬЮСИ**

*E.P. Istomin, V.V. Novikov, O.N. Kolbina, A.Yu. Sidorenko, S.Yu. Stepanov*

## **COMPLICATED INFORMATION SYSTEM RISK PREDICTION USING THE FILTER KALMAN–BUCY**

*Статья посвящена созданию нового подхода к организации разработки информационной системы прогнозирования рисков чрезвычайных ситуаций природного характера, обрабатывающей больше количество гетерогенной информации. Описан метод прогнозирования концентрации плотности льда, основанный на реализации фильтра Калмана–Бьюси в среде программы математического моделирования MATLAB с использованием обработанных данных, полученных со спутников над территорией Северного полюса.*

*Ключевые слова: информационная система, гетерогенные данные, управление территорией, природные процессы, фильтр Калмана–Бьюси.*

*This paper is devoted to creating a new approach to the development of information systems risk forecasting natural emergencies, handle a large number of heterogeneous information. Describes a method for predicting the density of ice concentration, based on the implementation of the filter Kalman–Bucy among mathematical modeling program MATLAB, using processed data from satellites over the territory of the North Pole.*

*Key words: information system, heterogeneous data, territory management, natural processes, the filter Kalman–Bucy.*

Интеграция информационных ресурсов в единую информационную среду и организация доступа к вычислительным ресурсам является одним из важнейших направлений развития современных информационных технологий. Проблемы создания и интеграции информационных ресурсов и продуктов являются приоритетными направлениями развития многих стран, в том числе и России, где эти проблемы отнесены к критическим направлениям развития современного общества. Тесное кооперирование информационных технологий и наук о Земле способствует пониманию как глобальных, так и региональных природных процессов, взаимодействия процессов, формирующих природную окружающую среду.

Следует отметить, что в настоящее время сбор данных об окружающей среде и корректное использование полученных таким образом наборов данных приобретает все большее значение. Данные об окружающей среде часто незаменимы, они всегда уникальны, хотя бы по времени их получения. Их сбор часто обходится очень дорого, в частности, который осуществляется по средствам космических спутников. По этим причинам огромную важность имеет задача извлечения максимальной пользы от данных, полученных в результате каждого исследования.

К данным об окружающей среде можно отнести такие природные характеристики, как температура, количество осадков, уровень и расход воды, скорость ветра, продолжительность процессов и др. При рассмотрении данных характеристик на протяжении многих лет выделяют их усреднённые и мгновенные значения. Если в пределах года определить экстремальные значения параметров тепла и увлажнённости, то процесс многолетних изменений можно будет представить как совокупность двусторонних фигур. Начало каждой из таких фигур связано с минимальным значением, которое сменяет через некоторое время максимальная величина. После максимума происходит спад к следующему минимуму, своеобразная кривая. Экстремальные ординаты разных по продолжительности двусторонних фигур представляют собой совокупности минимальных и максимальных значений природных характеристик [1].

Кроме термина «значение» как элемент многолетнего ряда наблюдений используют также понятие «событие». Событием называют обычно величины, превышающие или, наоборот, не превышающие некоторый критический уровень. Для выявления перехода значения многолетнего ряда природной характеристики в событие можно использовать статистические критерии или метод выделения уровней иерархии. При статистических критериях ориентируются на статистические параметры, среднее значение ряда и стандартное отклонение либо на величину заданной вероятности превышения. К примеру, превышение верхних критических уровней воды на реках приводит к наводнению.

Природные явления отличаются друг от друга, поэтому и для оценки событий используют различные критерии. Для паводков, половодий применяют отметку опасности экстремального явления. При анализе температур в качестве критерия оценки события применяют статистическую характеристику.

Для использования, а именно оценки и прогнозирования природных событий необходимо систематизировать данные о них. Самым верным решением будет создание информационной системы, предоставляющей полную и систематизированную информацию о различных аспектах природных явлений с одновременным раскрытием методики выбора приемлемых сценариев развития, т.е. возможностью прогнозировать риски появления природных чрезвычайных ситуаций.

Трудность при создании такой информационной системы связана с получением информации из различных источников. Исторические свидетельства находятся в архивах и библиотеках. Гидрометеорологические данные берутся из специализированных фондов и служб. В настоящее время к источникам информации о состоянии окружающей среды добавилось дистанционное аэро- и космическое зондирование больших участков поверхности Земли.

Необходимость интеграции разнородной информации об объектах окружающей среды требует развития ГИС в направлении расширения возможностей хранения информации в едином информационном пространстве. Данное требование может быть достигнуто в результате создания модели и методики организации геоинформационной системы управления территорией на основе гетерогенных баз данных.

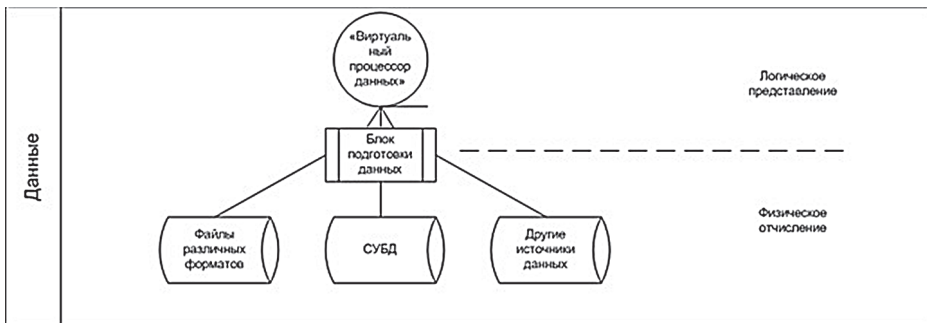
Данные об окружающем нас мире собраны в различных источниках и представлены в различных форматах, задача данной информационной системы собрать их воедино, не лишая возможности специалистов продолжать работать с ними на местах их сбора. Это возможно благодаря новым информационным технологиям, которые

могут предоставить разработчикам все возможные виртуальные ресурсы. Данная система должна успешно работать в многопользовательском режиме и предотвращать рассогласование данных, вызванное одновременной работой с одними и теми же данными, а также организовывать программную архитектуру.

Клиент-серверная модель соединений, взявшая за основу принципы структурного программирования, а именно разделения процессов ввода-вывода и вычислительной обработки, в последнее время не справляется с растущими потребностями разработчиков в области надежности и производительности систем.

Наиболее отвечающая всем требованиям равномерного распределения вычислительной нагрузки, обеспечения надежности и масштабируемости, является многоуровневая модель создания сложных программных систем. Она содержит три уровня представления данных «данные — обработка — интерфейс». Достоинствами данного представления являются независимость системы от количества программных компонентов, входящих в каждый уровень, свой набор протоколов и прикладных интерфейсов для каждого уровня, определяющий свою архитектуру.

Представления данных в такой информационной системе будет заключаться в создании «виртуального процессора данных», где будет проходить первичная обработка, приведение к единому формату и выявление связей между запрашиваемыми пользователем данными (рисунок).



Структурная модель представления данных в геоинформационной распределенной системе на основе гетерогенных баз данных

«Виртуальный процессор данных» определяется как программно-технологический модуль, который позволяет единообразно получить доступ ко всем данным системы через стандартные интерфейсы (протоколы) и скрывающий при этом особенности размещения данных, структуры и форматы данных. При этом нет необходимости преобразования данных к формату системы, они остаются в исходном представлении и могут использоваться локально своими программными продуктами.

«Блок подготовки данных» является промежуточным звеном между логическим представлением данных и их физическим расположением. В нем содержатся инструменты, протоколы, интерфейсы, позволяющие получить доступ к различным форматам данных и извлечь из файлов запрашиваемую информацию.

В информационной системе, разработанной с применением данной модели, предусмотрена работа с гетерогенными базами данных, с файлами и каталогами, а также другими источниками информации, которые можно обозначить как спутниковые снимки, файлы в форматах распространенных ГИС.

Помимо уникальности работы с данными геоинформационная система управления территориями должна помогать в принятии управленческих решений, что означает возможность прогнозирования появления чрезвычайных ситуаций природного характера и связанные с их появлением риски. Для этой задачи можно использовать различные статистические и математические методы обработки информации. Рассмотрим процесс применения математических методов для разработки метода прогноза плотности льда, относительно полных априорных данных анализа текущего состояния концентрации льда на тематическом слое северного полюса за наблюдаемый период времени. Данные были получены путем спутникового зондирования Земли и предоставлены лабораторией спутниковой океанографии, расположенной в Российском гидрометеорологическом университете, в городе Санкт-Петербург.

Алгоритмы, позволяющие решать задачи прогнозирования состояний при полной определенности исходных значений, множество (методы математической статистики, метод наименьших квадратов, максимального правдоподобия и другие). Одним из таких методов являются оптимальные фильтры [2]. Задачей фильтров является построение алгоритма, который осуществляет необходимое преобразование сигналов, поступающих на его вход. Наиболее универсальным фильтром для решения поставленной задачи является фильтр Калмана–Бьюси [3]. Главная особенность фильтра в том, что он обновляет оценку текущего состояния системы при поступлении новых данных, а затем предсказывает их будущее состояние. Сам же фильтр Калмана–Бьюси является рекуррентным алгоритмом, что дает преимущество реализации его на персональном компьютере.

Прогнозирование с помощью фильтра Калмана–Бьюси возможно только в случае, если прогнозируемые процессы в дискретном случае можно описать разностными уравнениями типа:

$$x_t = \Phi(t, t-1)x_{t-1}, \quad (1)$$

где  $\Phi(t, t-1)$  — переходная матрица, характеризующая структуру  $x_t$ , при этом наблюдения должны удовлетворять соотношению:

$$z_t = P_t x_t + \varepsilon_t, \quad (2)$$

где  $P_t$  — матрица ограничений на наблюдения;  $\varepsilon_t$  — ошибка наблюдений.

Для того чтобы решить задачу прогнозирования  $y(t)$  с помощью фильтра Калмана–Бьюси, необходимо провести зависимость:

$$\tilde{y}_i(t) = \sum_{j=0}^m a_{ij} \varphi_j(t), \quad i = \overline{0, n}, \quad (3)$$

где  $a_{ij}$  — случайные величины;  $\{\varphi_j(t)\}_{j=0}^m$  — непрерывные детерминированные функции времени в соответствие с выражением (1).

Требуемые преобразования можно осуществить на основе расширения пространства состояний [4], т.е. путем введения новых координат в пространство параметров, определяющих техническое состояние. Компонентами вектора  $x_t$  будут служить коэффициенты  $a_{ij}$  (элементы матрицы  $A$  выражения (3)), откуда  $x_t$  можно сформировать на строк матрицы  $A$ , где его размерность будет  $l = nm$ , где  $n$  и  $m$  — число строк и столбцов матрицы. При этом  $\Phi(t, t-1)$  — единичная матрица размером  $l \times l$ , а матрица  $P_t$  размера  $n \times l$  состоит из компонентов вектора  $\{\varphi_j(t)\}_{j=0}^m$  — детерминированного базиса модели (3).

Построение  $P_t$  можно реализовать последовательным сдвигом по строкам на  $i(m+j)$  транспонированного вектора  $\varphi(i - \text{номер строки матрицы } P_t)$  [5].

Решением уравнений (1) и (2) будут известные соотношения, определяющие фильтр Калмана–Бьюси [6]. После преобразования этих соотношений в соответствии с постановкой задачи прогнозирования получим:

$$x_{t+1} = x_t + W_t(z_t - P_t x_t), \quad (4)$$

$$W_t = K_t P_t^T (P_t K_t P_t^T + R_t)^{-1},$$

$$K_{t+1} = K_t - W_t P_t K_t, \quad (5)$$

где  $x_{t+1}$  — прогнозируемая на момент времени  $t+1$  по результатам имеющихся к этому времени наблюдений оценка  $x$  (условное математическое ожидание  $x$ );  $x_t$  — прогнозируемая на момент времени  $t$  оценка  $x$ ;  $K_t$  — ковариационная матрица компонентов вектора  $x_t$ ;  $K_{t+1}$  — ковариационная матрица компонентов вектора  $x_{t+1}$ ;  $W_t$  — матрица передачи фильтра;  $z_t$  — результаты наблюдений  $y(t)$  в момент  $t$ ;  $R_t$  — ковариационная матрица ошибок измерений  $\varepsilon_t$ .

Для прогнозирования, т.е. для решения выше приведенных уравнений, служат априорные оценки математических ожиданий — коэффициенты  $a_{ij}$ , их ковариационная матрица, результаты контроля  $y(t)$  и ковариационная матрица ошибок измерений. Следует отметить, что безотносительно к форме распределения входного сигнала при заданном первом и втором моментах фильтр Калмана–Бьюси обеспечивает получение наилучших в среднеквадратическом смысле линейных оценок.

Решением поставленной задачи будем использовать модель:

$$y(t) = a_0 + a_1 t + a_2 t^2, \quad (6)$$

так как выше сказано, что фильтр использует рекуррентную обработку информации, наделяя данный метод адаптивными свойствами, в свою очередь, позволяющими применять более простые модели  $y(t)$ , в частности модель вида:

$$y(t) = A \cdot F(t), \quad (7)$$

не выше второго порядка ( $m = 2$ ).

Данный метод был воспроизведен на основе данных плотности льда на протяжении Северного морского пути с помощью пакета программ математического моделирования MATLAB, в основе которого реализован фильтр Калмана–Бьюси.

Исходя из проведенного исследования, можно сказать, что при достаточном и достоверном объёме данных, примененный метод имеет положительный характер для дальнейшего более глубоко рассмотрения. Полученные результаты уже сейчас позволяют оказать помощь в научных экспедициях на Северный полюс как со стороны эффективности (до 30 %), так и со стороны безопасности. А применение данного метода в совокупности с геоинформационной системой управления территорией на основе гетерогенных баз данных, в которой доступна информация по множеству критериев представления окружающего пространства, может дать человечеству мощный инструмент борьбы со стихийными бедствиями и их разрушительной силой.

### Литература

1. *Елохин В.Р., Иванько Я.М., Старкова Н.В.* Модель информационной системы оценки изменчивости экстремальных природных явлений. // Известия ИГЭА, 2009, № 5.  
*Elokhin V.R., Ivanko Ya.M., Starkova N.V.* Model informatsionnoy sistemy otsenki izmenchivosti ekstremalnykh prirodnykh yavleniy. // Izvestiya IGEA, 2009, № 5.
2. *Васильев Б.В.* Прогнозирование надежности и эффективности радиоэлектронных устройств. — М.: Сов. радио, 1970. — 336 с.  
*Vasilyev B.V.* Prognozirovaniye nadezhnosti i effektivnosti radioelektronnykh ustroystv. — М.: Sov. radio, 1970. — 336 s.
3. *Браммер К., Зиффлинг Г.* Фильтр Калмана–Бьюси. — М.: Наука, 1982. — 200 с.  
*Brammer K., Ziffing G.* Filtr Kalmana–Byusi. — М.: Nauka, 1982. — 200 s.
4. *Мудров В.И., Кушко В.М.* Методы обработки измерений. — М.: Сов. радио, 1976. — 192 с.  
*Mudrov V.I., Kushko V.M.* Metody obrabotki izmereniy. — М.: Sov. radio, 1976. — 192 s.
5. *Абрамов О.В., Розенбаум А.Н.* Прогнозирование состояний технических систем. — М.: Наука 1990. — 20 с.  
*Abramov O.V., Rozenbaum A.N.* Prognozirovaniye sostoyaniy tekhnicheskikh sistem. — М.: Nauka 1990. — 20 s.
6. *Пугачев В.С.* Теория случайных функций и ее применение к задачам автоматического управления. — М.: Физматгиз, 1962. — 166 с.  
*Pugachev V.S.* Teoriya sluchaynykh funktsiy i yeye primeneniye k zadacham avtomaticheskogo upravleniya. — М.: Fizmatgiz, 1962. — 166 s.