

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 551.50 : 631.6

ПРИНЦИПЫ ФОРМИРОВАНИЯ СТРУКТУР ДАННЫХ ХРАНЕНИЯ, ОБРАБОТКИ И ПРЕДСТАВЛЕНИЯ ВРЕМЕННЫХ РЯДОВ МНОГОЛЕТНЕГО МОНИТОРИНГА

Н.К. Вахонин, кандидат технических наук

РУП «Институт мелиорации»

г. Минск, Беларусь

Ключевые слова: данные мониторинга, информационная система, временная ось, n-мерный гиперкуб, статистики, таблицы, ряды, колонки.

Введение

Принятие экономико-экологически эффективных решений по мелиорации, сельскохозяйственному использованию земель базируется на надежности зависимостей, описывающих функционирование рассматриваемых систем и их информационного обеспечения, получаемых на основе результатов мониторинга [1, 2].

Исходя из того, что любые реальные системы (природные, технические, агроприродные и т.п.) существуют в четырехмерном пространстве-времени, сформулирована геореляционная пространственно-временная концепция информационной системы хранения данных их мониторинга, реализующей информационное обеспечение поддержки принятия решений в различных предметных областях [2, 3].

В соответствии с методами Data Mining [4, 5] данные измерений любого показателя могут быть представлены в виде гиперкуба, главными осями которого является [3]: вид измеряемого показателя (осадки, уровни грунтовых вод, испарение и т.п.); точка измерения, положение которой в пространстве однозначно определяется трехмерными координатами x , y , z , с присвоенным ей именем (агроводобалансовый участок №1, водпост Ясельда-Трухановичи, метеостанция Пружаны $h=2$ м и т.п.); время измерения.

В данной работе рассматриваются особенности представления данных мониторинга в связи с многомерностью оси времени при иерархическом ее представлении календарной шкалой дат: год-месяц-число-час.

Обсуждение результатов

Все процессы в любых системах (природных, технических, природно-технических и т.д.) имеющие и непрерывный, и дискретный характер протекают во времени.

При этом очевидно, что характеристика любого процесса является тем более точной, чем с большей частотой (в идеале непрерывно) осуществляются измерения представляющей его переменной состояния. Соответственно частота измерения влияет и на точность установления влияния рассматриваемого процесса на другие процессы. К при-

меру фиксации изменения одной из определяющих величину формируемой урожайности переменных – значений температуры воздуха (T^o), со все большей частотой в течение периода вегетации (T_{veg}), дает основание надеяться получить (при достаточно продолжительных годах наблюдений) все более точную аппроксимацию строгой, имеющей место при непрерывном учете времени, зависимости (1), приближенной эмпирической зависимостью (2) при учете температур в дискретные моменты времени (t_i) периода вегетации:

$$\dot{o} = \int_{t=0}^{t=T_{veg}} f [T^o(t)] \times dt \quad (1)$$

$$\dot{o} = f_0(T_{t_0}^o) + f_1(T_{t_1}^o) + \dots + f_n(T_{t_{veg}}^o) \quad (2)$$

Однако так как проведение любых измерений имеет конечную скорость осуществления и ограничение по приемлемым затратам, то фиксация любой непрерывно изменяющейся переменной, неизбежно осуществляется с конечной частотой – в принятые с некоторым шагом дискретные моменты времени. Аналогично и при расчете процессов на реализуемых численными методами прогнозных моделях – переменные процессов также рассчитываются по дискретным временным шагам.

Очевидно, что хранение и обработка огромных массивов данных мониторинга вручную на бумажных носителях лишена смысла и должна осуществляться в компьютерном виде, объем памяти и быстродействие, которых, несмотря на постоянный рост, все же имеют ограниченный характер. В результате частота осуществления измерений (для быстроизменяющихся процессов решающим образом определяющая размеры массива наблюдений), помимо затрат на осуществление, лимитируется также памятью и быстродействием используемой компьютерной техники, определяющими возможность хранения и обработки собранных объемов данных.

С учетом вышеприведенных критериев выбор конкретных дискретов времени измерения целесообразно осуществлять в соответствии с характерным временем изменения переменных состояния каждого конкретного процесса, а при наблюдении переменных состояния процессов с различными характерными временами протекания, дискреты времени для каждого из них целесообразно принимать кратными для обеспечения измерений в одни моменты времени.

Очевидно, что на протяжении промежутка времени малого изменения переменной более частые измерения и хранение полученных данных практически бесполезны.

Среди определяющих формирование урожайности сельскохозяйственных культур переменных, наиболее быстроизменяющимися являются агрометеорологические показатели: температура, солнечная радиация, скорость ветра, имеющих характерное время изменения часы. При часовом дискрете число наблюдений за сутки составляет 24, а за

високосный год 8784. На практике, в соответствии с наставлениями гидрометслужбы, наблюдения осуществляются в 0, 3, 6, 9, 12, 15, 18, 21 часов, всего 8 измерений в сутки. Характерным временем изменения уровней грунтовых вод являются 5-10 дневки.

Таким образом, с практической точки зрения для реальных задач сельского хозяйства, мелиорации, экологии и т.п. более частые поминутные и посекундные наблюдения лишены практического смысла и дальнейшее рассмотрение осуществляется начиная с уровня срочных часовых наблюдений (при этом полученные выводы естественным образом распространяются и на наблюдения с любыми более частыми минутными, секундными и т.д. промежутками фиксации).

При характеристике процессов понятие времени имеет 2 вида: время момент и время период. Моменты срочных измерений переменных процессов (а также прогнозных значений, рассчитываемых на модели по временным шагам) характеризуются временем-моментом.

При календарном представлении времени одномерная временная ось заменяется многомерной иерархической шкалой, в каждом уровне вложенности которой отсчет времени начинается с начала: месяцев в каждом году, дней в каждом месяце, часов в каждом дне и т.д. При рассмотрении процессов с характерным временем изменения часы, время-момент осуществления измерения в календарном виде представляется четырехмерным вектором: датой в виде год-месяц-день-час. Таким образом измерения данного вида переменной в заданном месте является точкой в четырехмерном временном пространстве, являющемся эквивалентом одномерной оси времени.

Однако возможен и другой вариант отсчета времени: в каждом году, месяце, сутках оно может осуществляться и по одномерной шкале с принятой минимальной ценой деления, в частности в рассматриваемом случае в часах: от 1 до 8784 часа в году; от 1 до 744 часа в месяце, от 1 до 24 часа в сутках.

Рассчитываемые при агрегировании во времени статистические характеристики рядов данных исходных срочных наблюдений характеризуются временем-периодом, в общем случае представляемым двумя датами: начала и конца периода агрегирования (при кусочном его характере – набором пар дат или полным их перечислением).

Следует отметить, что естественно-физически выделяющимися промежутками времени являются сутки (дни) и годы, определяемые периодом обращения Земли соответственно вокруг своей оси и вокруг Солнца. Часы и месяцы представляют собой формальное разбиение на равные периоды соответственно суток на 24 промежутка и года на 365 (366) промежутков.

Помимо этого на практике дополнительно используются чисто формально выделяемые периоды: декады (36 десятидневок в году) и, имеющие природно-физический смысл, 4 сезона года (зима, весна, лето, осень). Возможно также подразделение суток на более детальные периоды: светлое-темное время или еще детальнее утро, день, вечер, ночь.

Для процессов с большими характерными временами изменения переменных (месяц, сезон, год) и само срочное наблюдение часто по умолчанию относят к периоду в целом, пренебрегая указанием конкретного момента (даты) его осуществления. Однако для обеспечения однородности данных измерения следует осуществлять в одну дату с хранением ее наряду с результатом замера в базе данных. В этом случае единственное измерение одновременно является и всеми статистиками: средним, максимальным, минимальным значением за рассматриваемый период.

Следует также отметить, что наряду с вышеприведенными календарно неизменными периодами в различных предметных областях используются имеющие физический смысл периоды со скользящими по годам датами начала и конца: предпосевной период, период половодья, период вегетации, меженный период в гидрологии и растениеводстве и т.п. для которых также могут считаться статистики по исходным срочным замерам.

Более строгий анализ происходящих изменений наблюдаемых процессов (выявление трендов, периодичностей и т.п.), получение регрессионных зависимостей других переменных систем, зависящих от наблюдаемых переменных, имеет место при рассмотрении наиболее детальных данных – срочных часовых наблюдений, хранение и обработка которых и должна обеспечиваться в информационной системе мониторинга.

Однако в ряде задач могут потребоваться не исходные наблюдения, а генерализованные до все более крупных периодов. К примеру при получении ранее приведенной регрессионной зависимости урожайности от температур воздуха, очевидно более строгой является аппроксимация полиномом с использованием почасовых значений. Однако для надежной идентификации большого числа коэффициентов, в этом случае требуются чрезвычайно продолжительные ряды наблюдений. При их отсутствии вынужденно приходится использовать в качестве аргументов в зависимости урожая, данные за последовательно все более крупные промежутки времени (дни, декады, месяцы и т.п.). При этом очевидно, что с увеличением периода учета данных последовательно уменьшаются основания на получение надежной зависимости урожайности.

Иерархичность периодов измерений приводит к тому, что интерес в различных задачах, могут представлять статистики, посчитанные по срочным часовым наблюдениям для суточных, декадных, месячных, годовых и многолетних периодов, а также статистики от суточных статистик, и далее статистики посчитанные по всем вышеперечисленным периодом для многолетия.

Используемая многомерная иерархическая шкала вложенных периодов времени порождает при решении различных задач потребность последовательного агрегирования данных исходных замеров – расчета по ним различных статистик для все более крупных временных периодов.

Иерархии всех возможных видов статистик приведены на рис. 1.

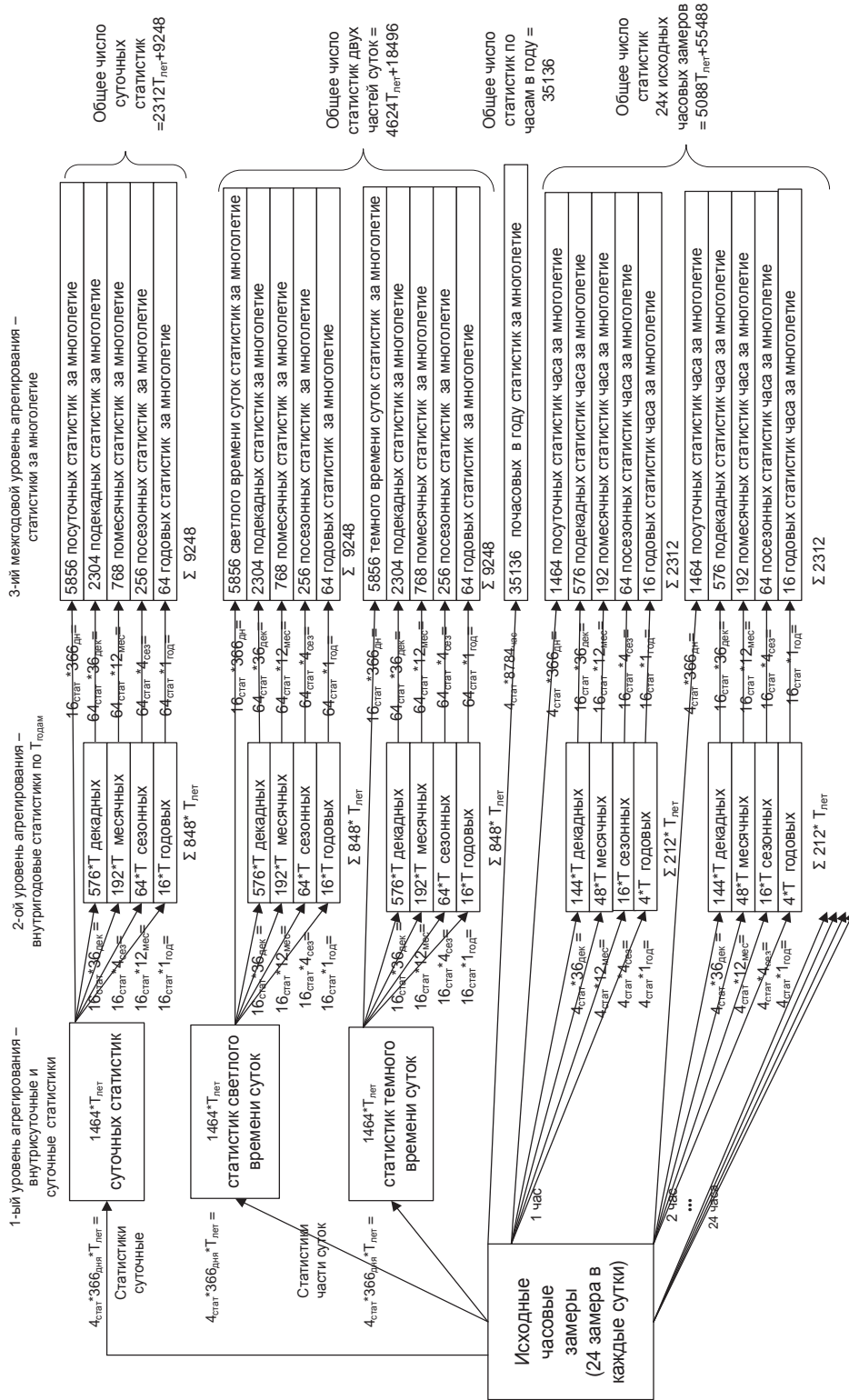


Рисунок 1 – Все варианты агрегирования данных мониторинга во времени (4 типа статистик: max, min, сред., дисперсия) с расчетом всех сочетаний статистик (max min, min max, max min max и т.п.) на каждом уровне агрегирования

При этом имеет место формирование статистик трехуровневой вложенности (рис. 1):

1 уровень – суточные и внутрисуточные статистики от срочных часовых наблюдений, рассчитываемые по исходным, соответственно по 24 и 12 значениям измерений при ежечасных наблюдениях.

2 уровень – внутригодовые статистики: суточные (366 суток), декадные (36 декад), месячные (12 месяцев), сезонные (4 сезона), годовые (1 год), рассчитываемые для каждого из 24х часов исходных наблюдений, а также из внутрисуточных статистик – 2х (светлое, темное время суток) или 4х (при разбиении суток на утро, день, вечер и ночь) и из суточных статистик для каждого из 365 (366) дней года.

3 уровень – многолетние (межгодовые) статистики: часовые, в сутках, часовые в году, суточные, декадные, месячные, сезонные, годовые статистики, рассчитываемые от соответствующих статистик годового уровня, а также суточного уровня и исходных часовых данных.

Расчет каждой из месячных, сезонных и годовых статистик 2-го и 3-го уровня агрегирования может осуществляться как непосредственно от исходных часовых замеров или суточных статистик (в соответствии с последовательностью, приведенной на рис.1), так и последовательно используя соответствующие, уже рассчитанные статистики данного уровня агрегирования для более крупных периодов: декадные для расчета месячных, месячные для сезонных, сезонные для годовых, что значительно сокращает объем вычислений.

Аналогично этому по различному могут рассчитываться многолетние статистики 3-го уровня агрегирования для различных все более крупных периодов (дней – декад – месяцев – сезонов – годов).

На рис. 1 приведены все варианты статистик, посчитанные для случая 4 основных базовых статистик: max, min, сред., дисперсия, причем с полным их сочетанием между собой на каждом уровне агрегирования, т.е. с расчетом $max_{дек} min_{сут}$; $min_{дек} max_{сут}$; $max_{многол}$, $min_{мес}$, $сред_{сут}$ и т.д.

Для общности дополнительно рассчитано также количество статистик для частей суток.

Общее количество вычисляемых значений статистик, представленных на рис.1, составляет $12024 * T_{лет} + 118368$ значений (где $T_{лет}$ – число лет наблюдений).

Более сокращенный вариант – расчет на разных уровнях агрегирования статистик только аналогичных их виду с прошлого уровня: $сред_{мес}$ $сред_{сут}$; $max_{многол}$ $max_{мес}$ $max_{сут}$ и т.п., требует вычисления $10126T_{лет} + 80388$ значений статистик (рис. 2).

Очевидно, что во всех случаях число статистик растет линейно с увеличением лет наблюдений. Увеличение числа рассматриваемых базовых статистик ведет к геометрическому росту вычисляемых статистик в соответствии с рис.1, 2, что требует значительных вычислительных ресурсов, как памяти, так и быстродействия.

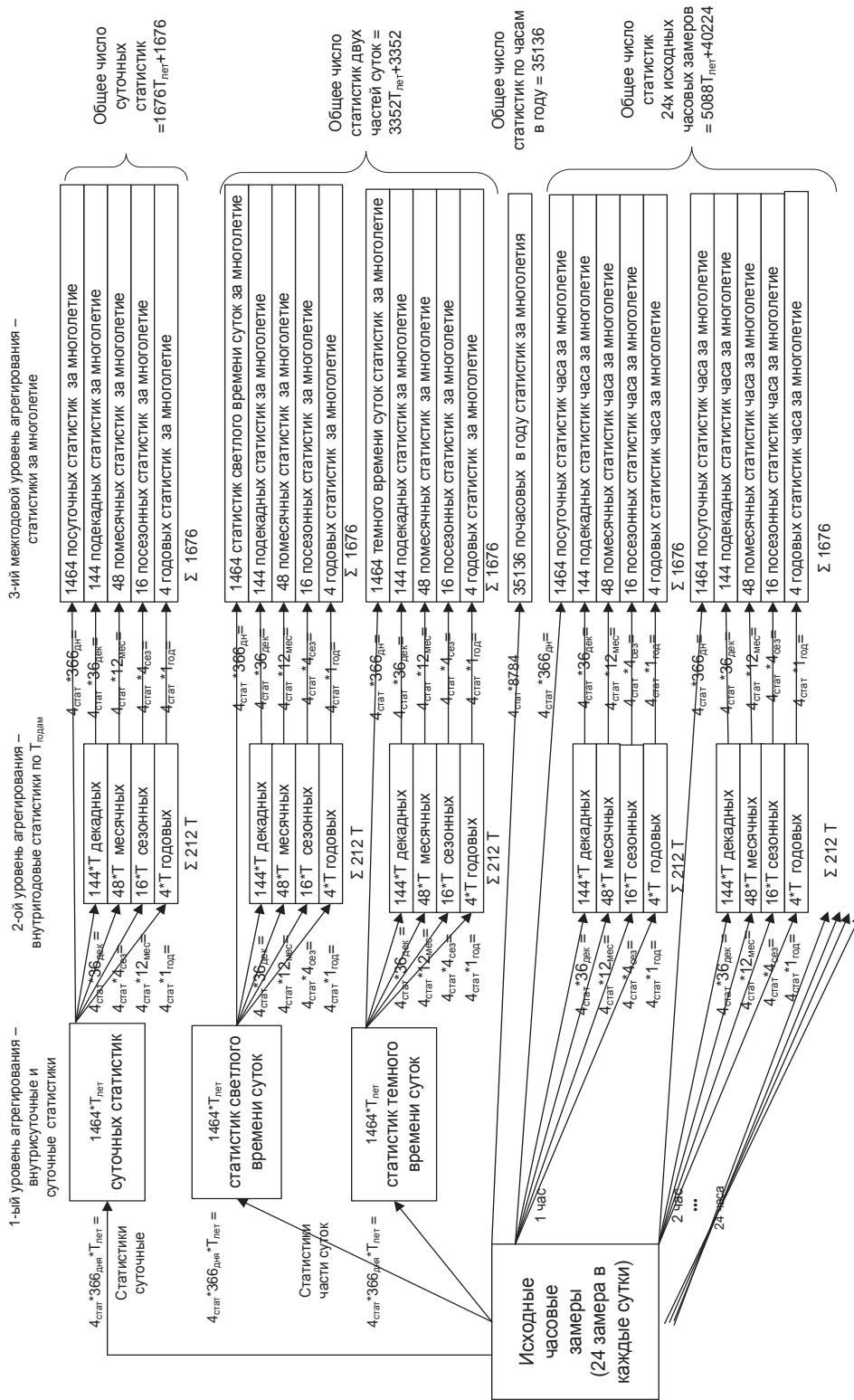


Рисунок 2 – Все варианты агрегирования данных мониторинга во времени (4 типа статистик: max, min, сред., дисперсия) при расчете только однотипных статистик на каждом уровне агрегирования

Для реально требуемых и используемых представлений результатов многолетних измерений по различным периодам, представляющих гиперкуб (гиперматрицу) в n-мерном временном пространстве с временными осями: часы, дни, декады, месяцы, сезоны, годы, многолетия, необходимо его проецирование на различные плоские (двухмерные, одномерные) таблицы у которых обе оси – различные сочетания пар всех видов временных периодов.

Так как в реляционных базах данных данные хранятся в форме двумерных атомарных таблиц, то при выделении одного измерения под время в виде полной даты измерения, представляемой сложным именем год-месяц-день-час (при необходимости беспроблемно детализируемым до минут-секунд и т.д.), второе измерение остается под результатами замеров. В результате для каждого вида наблюдения, осуществляемого в конкретной пространственной точке, таблица базы данных наблюдений представляется двойкой: дата измерения-результат замера.

При этом формы ввода данных, структуры данных формируемые из базы данных для различных последующих обработок, а также создаваемые выходные отчеты могут представляться двумерными и одномерными таблицами различного вида.

Ввод годовых данных срочных часовых замеров по каждому из 24 часов суток за 365 (366) суток в году может обеспечиваться через любой из шести альтернативных вариантов таблиц (табл.1-3), представленных тремя парами таблиц одного содержания двухмерных и одномерных видов.

Помимо вида (1)-(3), (1')-(3') все 6 вариантов таблиц могут использоваться и в транспонированном виде. При этом следует отметить, что в некоторых информационных системах ограничение по возможному числу колонок обычно намного более жесткие, чем по числу строк. Поэтому, к примеру в Excel, ограниченном 256 колонками, таблицы (1')-(3') в транспонированном виде не могут быть размещены на странице, в отличие от таблиц исходного вида (табл.(1), (3)). Таким образом правильный выбор варианта транспонирования обеспечивает возможность размещения таблицы с одним и тем же составом информации, при критических условиях по возможному числу колонок в ней.

Для ввода всех годовых данных необходимо: 12 экземпляров таблиц вида №1 или №1', или 24 экземпляра на год таблиц вида №2 или №2', или по одной на год таблиц №3 или №3'.

Учитывая привычность календарного характера восприятия времени, более удобными для первичного ручного ввода данных в базу данных являются двумерные таблицы: табл. 1 (12 на год по числу месяцев) и табл. 2 (24 на год при ежечасных наблюдениях в сутках, или 8 – при обычно имеющихся наблюдениях с трехчасовым периодом).

Следует отметить, что таблицы (1) – (3) помимо данных исходных замеров позволяют рассчитывать и хранить соответствующие статистики по строкам и столбцам (справа и снизу таблиц).

Таблица 1 – Месячная таблица срочных часовых замеров двухмерная 2015 г., январь

часы суток \ дни месяца	1	2	...	23	24
1					
...					
31					

Таблица 1' – Месячная таблица срочных часовых замеров одномерная 2015 г., январь

часы месяца	замер
1	
2	
...	
744	

Таблица 2 – Годовая таблица срочных замеров на конкретный час или одной из суточных статистик двухмерная 2015 г., 9.00 час.

дни месяца \ месяцы	1	2	...	12
1				
...				
31				

Таблица 2' – Годовая таблица срочных замеров на конкретный час или одной из суточных статистик одномерная 2015 г., 9.00 час.

дни года	замер
1	
2	
...	
366	

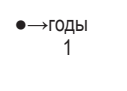

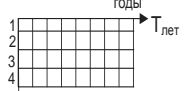
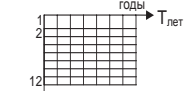
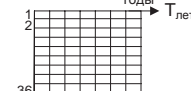


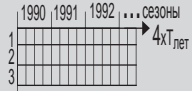
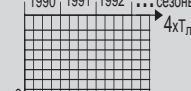
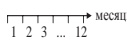
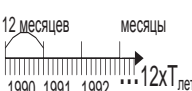
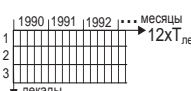
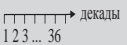

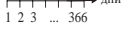
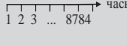
Таблица 3 – Годовая таблица срочных замеров на все часы двухмерная 2015 г., январь

дни суток \ часы	1	2	...	23	24
1					
2					
...					
366					

Таблица 3' – Годовая таблица срочных замеров на все часы одномерная 2015 г., январь

часы года	замер
1	
2	
...	
8784	

Таблица 4 – Все типы таблиц вариантов двухмерной временной развертки данных (4 основные статистики) конкретного вида

период	многолетие	годы	сезоны	месяцы	декады
Многолетие в годах					
ГОДЫ	 <p>● → годы 1</p> <p>112табл×4стат =448табл</p>	 <p>1 год годы</p> <p>таблиц годовых – 112 рядов в годов. табл. 112</p>	 <p>годы → T_{лет}</p> <p>1 2 3 4 ↓ сезоны</p> <p>таблиц сезонных – 112 рядов в сезон. табл. 448</p>	 <p>годы → T_{лет}</p> <p>1 2 3 12 ↓ месяцы</p> <p>таблиц месячных – 112 рядов в месяч. табл. 1344</p>	 <p>годы → T_{лет}</p> <p>1 2 36 ↓ декады</p> <p>таблиц декадных – 112 рядов в декад. табл. 4032</p>
Многолетие в сезонах					
СЕЗОНЫ	 <p>1 2 3 4 → сезоны</p> <p>112табл×4стат =448табл</p>	 <p>4 сезона 1990 1991 1992 → 4хT_{лет}</p> <p>таблиц сезонных – 112 рядов в сезон. табл. 112</p>	 <p>1990 1991 1992 ... сезоны 1 2 3 → 4хT_{лет}</p> <p>↓ месяцы</p> <p>таблиц месячных – 112 рядов в месяч. табл. 336</p>	 <p>1990 1991 1992 ... сезоны 1 2 3 → 4хT_{лет}</p> <p>↓ декады</p> <p>таблиц декадных - 112 рядов в декадн. табл. 1008</p>	
Многолетие в месяцах					
МЕСЯЦЫ	 <p>1 2 3 ... 12 → месяцы</p> <p>112табл×4стат =448табл</p>		 <p>12 месяцев месяцы 1990 1991 1992 ... → 12хT_{лет}</p> <p>таблиц месячных – 112 рядов в месяч. табл 112</p>	 <p>1990 1991 1992 ... месяцы 1 2 3 → 12хT_{лет}</p> <p>↓ декады</p> <p>таблиц декадных - 112 рядов в декадн. табл 336</p>	
Многолетие в декадах					
ДЕКАДЫ	 <p>1 2 3 ... 36 → декады</p> <p>112табл×4стат =448табл</p>			 <p>36 декад декады 1990 1991 1992 ... → 36хT_{лет}</p> <p>таблиц декадных - 112 рядов в декадн.табл 112</p>	
Многолетие в днях					
СУТКИ	 <p>1 2 3 ... 366 → дни</p> <p>28табл×4стат= 112 табл</p>				
Многолетие в часах					
ЧАСЫ	 <p>1 2 3 ... 8784 → часы</p> <p>1табл×4стат= 4 табл</p>				
Σт.	Σ1908	112	224	336	448
Σр.		112	560	1792	5488

**многолетних наблюдений (годы по горизонтали (в колонках)) и их статистик
и точки осуществления мониторинга**


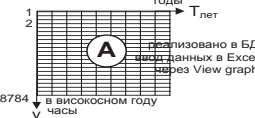
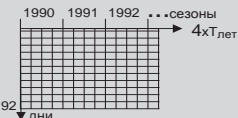
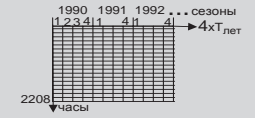
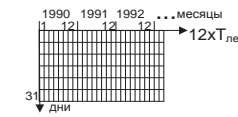
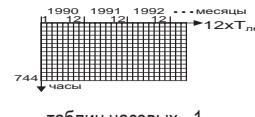

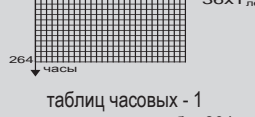
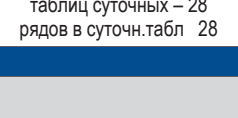
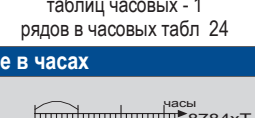




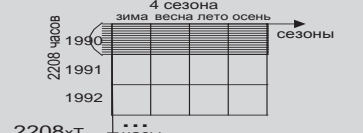
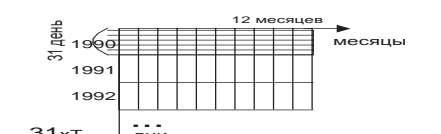
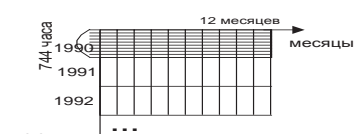
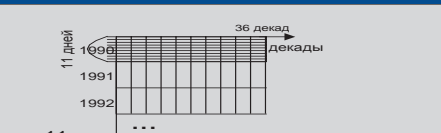
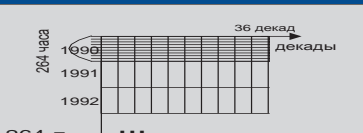
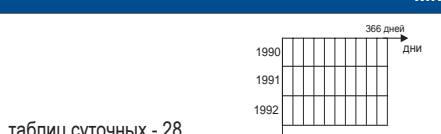
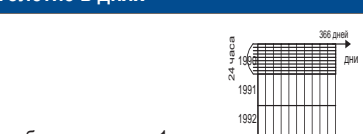

сутки исходные замеры и статистики (max, min, сред. и т.д.) суточного агрегирования часовых данных	часы (исходные срочные замеры)	Всего типов таблиц	Всего таблиц
Многолетие в годах			
 <p>таблиц суточных – 28 рядов в суточ.табл. 10248, в т.ч. (4стат×366дн=1464ряд) (остальные 24час.табл.×366=8784 ряда повторяют ряды табл.А)</p>	 <p>таблиц часовых - 1 рядов в часовых табл. 8784 колонок Tлет штук</p>	7	925
Многолетие в сезонах			
 <p>таблиц суточных – 28 рядов в суточн. табл. 2576</p>	 <p>таблиц часовых - 1 рядов в часовых табл 2208</p>	6	813
Многолетие в месяцах			
 <p>таблиц суточных – 28 рядов в суточн.табл. 868</p>	 <p>таблиц часовых - 1 рядов в часовых табл. 744</p>	5	701
Многолетие в декадах			
 <p>таблиц суточных – 28 рядов в суточн.табл 308</p>	 <p>таблиц часовых - 1 рядов в часовых табл 264</p>	4	589
Многолетие в днях			
 <p>таблиц суточных – 28 рядов в суточн.табл 28</p>	 <p>таблиц часовых - 1 рядов в часовых табл 24</p>	3	141
Многолетие в часах			
	 <p>таблиц часовых - 1 рядов в часовых табл. 1</p>	2	5
140	6	27	3174
14028	12025		

Таблица 5 – Все типы таблиц вариантов двумерной временной развертки данных (4 основных статистики) конкретного вида и точки

Пер	Многолетие	Годы	Сезоны	Месяцы	Декады
Многолетие в годах					
ГОДЫ	<p>112табл×4стат=448табл</p>	<p>таблиц годовых -112 рядов в годов.таб.112*Т_{лет}</p>	<p>таблиц сезонных – 112 рядов в сезон. табл. 448*Т_{лет}</p>	<p>таблиц месячных – 112 рядов в мес.таб.1344*Т_{лет}</p>	<p>таблиц декадных – 112 рядов в дек.табл.4032*Т_{лет}</p>
Многолетие в сезонах					
СЕЗОНЫ	<p>112табл×4стат=448табл</p>	<p>таблиц сезонных – 112 рядов в сезон. табл. 112*Т_{лет}</p>	<p>таблиц месячных – 112 рядов в мес.таб.336*Т_{лет}</p>	<p>таблиц декадных – 112 рядов в дек.табл.1008*Т_{лет}</p>	
Многолетие в месяцах					
МЕСЯЦЫ	<p>2табл×4стат=448табл</p>			<p>таблиц месячных – 112 рядов в мес.таб.112*Т_{лет}</p>	<p>таблиц декадных – 112 рядов в дек.табл.336*Т_{лет}</p>
Многолетие в декадах					
ДЕКАДЫ	<p>112табл×4стат=448табл</p>				<p>таблиц декадных – 112 рядов в дек.табл.112*Т_{лет}</p>
Многолетие в днях					
СУТКИ	<p>28табл×4стат=112 табл</p>				
Многолетие в часах					
ЧАСЫ	<p>1табл×4стат=4 табл</p>				
Σ табл	1908	112	224	336	448
Σ рядов		112*Т_{лет}	600*Т_{лет}	1792*Т_{лет}	5488*Т_{лет}

**многoletних наблюдений (годы по вертикали (в строках)) и их статистик
осуществления мониторинга**

Сутки исходные замеры и статистики (max, min, сред. и т.д.) суточного агрегирования часовых данных	Часы (исходные срочные замеры)	Всего типов таблиц	Всего таблиц
Многолетие в годах			
 <p>таблиц суточных - 28 рядов в суточн.табл. $10248 \times T_{\text{лет}}$</p>	 <p>таблиц часовых - 1 рядов в час.табл. 8784</p>	7	925
Многолетие в сезонах			
 <p>таблиц суточных - 28 рядов в суточн.табл. $2576 \times T_{\text{лет}}$</p>	 <p>таблиц часовых - 1 рядов в час.табл. $2208 \times T_{\text{лет}}$</p>	6	813
Многолетие в месяцах			
 <p>таблиц суточных - 28 рядов в суточн.табл. $868 \times T_{\text{лет}}$</p>	 <p>таблиц часовых - 1 рядов в час.табл. $744 \times T_{\text{лет}}$</p>	5	701
Многолетие в декадах			
 <p>таблиц суточных - 28 рядов в суточн.табл. $308 \times T_{\text{лет}}$</p>	 <p>таблиц часовых - 1 рядов в час.табл. $264 \times T_{\text{лет}}$</p>	4	589
Многолетие в днях			
 <p>таблиц суточных - 28 рядов в суточн.табл. $28 \times T_{\text{лет}}$</p>	 <p>таблиц часовых - 1 рядов в час.табл. $24 \times T_{\text{лет}}$</p>	3	141
Многолетие в часах			
	 <p>таблиц часовых - 1 рядов в час.табл. 1</p>	2	5
<p>140 $14028 \times T_{\text{лет}}$</p>	<p>6 $12025 \times T_{\text{лет}}$</p>	27	3174

При этом наиболее целесообразной (мнемоничной), с точки зрения обеспечения дружелюбного пользовательского интерфейса при вводе данных, представляется обычно используемая в бумажных журналах наблюдений в метеорологии, гидрологии и т.п. форма двумерной годовой таблицы вида 12 месяцев * 31 день (табл.2), по экземпляру для каждого часа наблюдений.

В различного вида задачах могут потребоваться формы таблиц для представления как исходных данных, так и для всех вышеприведенных на рис.1, рассчитанных по ним статистик.

Все виды плоских таблиц с осями времени и по горизонтали и по вертикали (т.е. и по столбцам, и по строкам), приведенные в таблицах 4,5, являются исчерпывающим перечнем форм, представляющих проекцию 7-мерного (включая многолетие) гиперкуба данных мониторинговых измерений на двумерные таблицы со всеми комбинациями пар временных осей, обеспечивают все многообразие возможных представлений данных многолетних измерений и любых их статистик.

В ячейках табл. 4, приведены различного вида двумерные таблицы данных мониторинга. В колонках (по горизонтали) таблиц представлены годы периода наблюдений, со все возрастающей сверху вниз по строкам табл.4 детализацией (ценой деления) шкалы времени внутри лет наблюдений, начиная от годов последовательно до сезонов, месяцев, декад, суток, часов.

Второе временное измерение в таблицах, приведенных в табл.4, представлено в их строках (по вертикали). При этом в таблицах расположенных справа-налево в колонках табл.4, вертикальная ось все более генерализована: часы-сутки-декады-месяцы-сезоны-годы-многолетия.

В табл.5 годы наоборот представлены по вертикали (по строкам) с дальнейшей детализацией, по смыслу аналогичному табл.4.

Все таблицы могут формироваться в том виде, как они представлены в табл. 4,5 и в транспонированном виде, выбирая конкретный вид с учетом лимитированности возможного числа колонок и строк таблиц в конкретной информационной системе.

Каждая из таблиц в табл.4, 5 может дополняться, справа и внизу статистиками рассчитываемыми по каждой из двух временных осей: соответственно по строкам и столбцам.

Формирование таблиц любого вида, представленных в табл.4, 5, может осуществляться небольшим набором процедур, используя на входе таблицу вида А – часовых данных за многолетие.

Заключение

Учитывая стохастичность протекающих в естественно-природной среде процессов, для получения статистически надежных выводов по оценке их изменения, получения регрессионных зависимостей с их использованием необходимо осуществление их многолетнего мониторинга.

Характерным временем протекания процессов в различных системах на макроуровне являются часы. В связи с этим, при проведении наблюдений представляющих их переменных, достаточно осуществление их с часовой частотой.

Для эффективного хранения, обработки и использования результатов мониторинга необходимо их формирование в правильно организованную, ориентированную на работу с временными рядами информационную систему.

При использовании календарного исчисления одномерная временная ось разбивается на 7 иерархически вложенных периодов.

Исходя из этого рассчитаны все варианты агрегирования во времени, причем в связи с естественно – физической иерархичностью времени выделено 3 уровня агрегирования: суточный, годовой, многолетний.

В результате проецирования 7-мерного гиперкуба на плоские таблицы с различными сочетаниями двух временных осей получены все возможные варианты представления данных многолетнего мониторинга в плоских таблицах. Рассчитанное число различных необходимых таблиц, рядов и колонок в них и рассмотрение возможности транспонирования, обеспечивает размещение таблиц в электронном виде при критичном числе колонок.

Проведенные выкладки являются концептуальной основой реализации базы данных временных рядов мониторинга, организации взаимодействия ее с электронными таблицами Excel, программами статистических расчетов и т.д.

Библиографический список

1. Вахонин, Н.К. Концептуальные принципы создания единой информационной системы поддержки принятия решений в мелиоративной отрасли / Н.К. Вахонин // Мелиорация. - №1(69), Мн., РУП "Институт мелиорации", 2013. – С.5-19.
2. Вахонин, Н.К. Принципы организации мониторинга для информационного обеспечения принятия решений в мелиорации/ Н.К. Вахонин// Мелиорация», №2(60), Мн., РУП "Институт мелиорации", 2008. – С.5-22.
3. Вахонин, Н.К. Теоретические основы построения информационной системы хранения данных мониторинга / Н.К. Вахонин // Достижения и перспективы инновационного развития мелиоративной науки Беларуси: доклады междунар. науч. конф. (Минск, 14-16 дек. 2010 г.). – С.84-88.
4. Michael J.A. Berry, Gordon Linoff. Data Mining Techniques – Jhon Wiley and Sons, 1997.
5. Барсегян, А.А. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, П.И. Холод // – БХВ-Петербург, 2004.

Summary

N. Vakhonin

PRINCIPLES OF DATA STRUCTURES, STORAGE, PROCESSING AND PRESENTATION OF TIME SERIES OF LONG-TERM MONITORING

The article describes the principles of information system for data long-term monitoring according to the hierarchy of the time axis in calendar time representation. All the statistics options at three levels of aggregation of multi-year term hours of observation time are considered. We present all the options projecting an n-dimensional hypercube time series of measurements on planar two-dimensional table calculating their size. The possibilities of placing the data in the tables at the critical number of columns, through their transposition are analyzed.

Поступила 21.10.2015