Интеграция временных рядов спутниковых и метеорологических данных со статистическими данными в информационных системах мониторинга

И. А. Уваров¹, А. Е. Платонов², А. В. Титков², В. А. Толпин¹, В. В. Малеев²

¹ *Институт космических исследований РАН, Москва, 117997, Россия*
E-mail: smis@smis.iki.rssi.ru

² *Центральный научно-исследовательский институт эпидемиологии
Роспотребнадзора, Москва, 111123, Россия*
E-mail: platonov@pcr.ru

Совместный анализ статистических данных, привязанных к единицам административно-территориального деления, и продуктов обработки спутниковой информации, а также метеорологических данных представляет интерес для выявления зависимостей, имеющихся между наблюдаемыми величинами. Интеграция статистических данных в информационные системы требует создания специальных средств автоматизированной обработки, архивации, разработки инструментов интерактивного анализа. Они обеспечивают приведение всех видов исходных данных к сопоставимому виду, агрегирование множества пространственно распределённых и повторяющихся во времени измерений к ряду интегральных характеристик. В работе представлены возможности решения данной задачи в рамках информационной системы «Вега-Science» (уникальная научная установка, входящая в Центр коллективного пользования «ИКИ-Мониторинг»). Приведены примеры использования созданных решений для анализа данных о состоянии сельскохозяйственной и лесной растительности, а также о распространении опасных инфекционных заболеваний.

Ключевые слова: дистанционный мониторинг, состояние растительного покрова, природно-очаговые инфекции, информационная система, метеорологические данные

Одобрена к печати: 08.12.2020

DOI: 10.21046/2070-7401-2020-17-7-105-114

Введение

Создание специализированных информационных систем мониторинга сопряжено со сбором и обработкой множества количественных характеристик, в первую очередь статистических, имеющих пространственную привязку к единицам административно-территориального деления и периодичность наблюдения через определённые временные интервалы. Примерами таких данных могут выступать показатели заболеваемости населения, в том числе от опасных инфекций (Прислегина и др., 2021), статистические показатели промышленного и сельскохозяйственного производства, работы транспорта и прочее. Значительная часть характеристик такого рода может влиять и/или зависеть от различных параметров состояния окружающей среды, например типа и состояния растительного покрова, уровня загрязнения, различных метеорологических параметров (Платонов и др., 2019). Многие из этих величин сегодня могут быть получены на основе спутниковых наблюдений (Platonov et al., 2019). В то же время метеоданные и продукты обработки спутниковых наблюдений представляют собой пространственно-временной массив данных, причём частота их получения (отличающаяся для разных видов данных) значительно превосходит частоту получения различной статистической информации. Поэтому при изучении зависимости различных процессов, описываемых статистическими данными, требуется проведение специальной обработки спутниковых и метеоданных и интеграция их на те же пространственные и временные интервалы, по которым имеется статистическая информация. Стоит отметить, что когда требуется установление зависимостей процессов, описываемых статистическими данными, особенно в случае необходимости выявления и обоснования достоверности долговременных трендов, необходима обработка значительных объёмов спутниковых и метеорологических данных и создание специализированных инструментов

для проведения анализа интегрированных на пространственные единицы (участки) и временные интервалы результатов обработки спутниковых и метеоданных. Для этого требуется:

- организация полностью автоматизированных процедур интеграции спутниковых и метеоданных;
- обеспечение автоматизированного ведения архивов интегрированных данных;
- обеспечение инструментов анализа данных (как пространственных, так и временных рядов).

В настоящей работе представлена система, реализованная на базе Центра коллективного пользования (ЦКП) «ИКИ-Мониторинг» (Лупян и др., 2019) и на основе метеоданных (Уваров и др., 2013), которая обеспечивает решение вышеуказанных задач. Приведены примеры использования системы вместе со статистическими данными.

Особенности организации статистических данных

Статистические данные, имеющие пространственную привязку к единицам административно-территориального деления и периодичность наблюдения через определённые временные интервалы, описывают широкий спектр природных, хозяйственных, социальных процессов и явлений. Далее подробнее рассмотрим работу с данными, в отношении которых имеется опыт интеграции в информационные системы мониторинга, в частности с показателями заболеваемости населения от опасных инфекций, а также численными характеристиками сельскохозяйственного производства.

Поступление статистических данных в информационную систему, выполняющую роль интегратора, обычно происходит в виде табличных файлов различных форматов. При этом может варьироваться структура таблиц в зависимости от количества обрабатываемых показателей. Общая черта всех структур данных такого рода заключается в наличии привязки ко времени и единицам административного деления.

Данные Центрального научно-исследовательского института эпидемиологии Роспотребнадзора характеризуют заболеваемость населения следующими видами инфекций: астраханская риккетсиозная лихорадка, крымская геморрагическая лихорадка, иксодовый клещевой боррелиоз, клещевой вирусный энцефалит, лихорадка Западного Нила, сибирский клещевой тиф, геморрагическая лихорадка с почечным синдромом.

Единица пространственной привязки таких данных — субъекты РФ. Часть информации также доступна по отдельным административным районам. Период наблюдения составляет один год. В разрезе принадлежности к территориальным единицам, периоду времени и типу инфекции выделяется несколько статистических показателей, в том числе:

- количество случаев заболевания и заболеваемость на 100 тыс. населения;
- количество случаев смертей и смертность на 100 тыс. населения;
- количество «утраченных лет полноценной жизни» за год в регионе в абсолютном выражении и на 100 тыс. населения;
- экономические потери за год в регионе в абсолютном выражении и в пересчёте на одного жителя;
- показатель заболеваемости сельских жителей по отношению к городским.

Анализ состояния сельскохозяйственной растительности обеспечивается использованием информации о фактической и ожидаемой урожайности, а также данных сельскохозяйственной переписи. Периодичность этих данных составляет один год. Ожидаемая урожайность относится к текущему году по итогам пройденного периода года.

Особенности организации спутниковых и метеоданных

Данные спутниковых наблюдений используются в информационных системах мониторинга в виде ряда продуктов обработки, таких как:

- композитные изображения, очищенные от влияния облачности, теней и шумов;
- вегетационные индексы;
- карты растительного покрова и др.

Периодичность получаемых продуктов обработки спутниковых данных значительным образом варьируется. Композитные изображения и вегетационные индексы могут быть получены на регулярной основе через промежутки времени от одного дня до целого сезона. Тематические карты, в том числе карты растительного покрова, строятся ежегодно.

Метеорологические данные включают данные прогноза и реанализа, полученные моделью Global Forecast System Национальных центров экологического прогнозирования США (англ. National Centers for Environmental Prediction — NCEP). Данные из различных массивов пространственно локализованы в узлах регулярной сетки с шагом 1; 0,5 или 0,25°. Периодичность поступления таких данных составляет 3 или 6 ч.

В наборах метеорологических данных представлено свыше 150 различных показателей, причём доступны как данные на поверхности земли, так и на высоте (на изобарических поверхностях) и в почве. Важнейшими показателями для исследования состояния окружающей среды выступают температура, количество осадков, влажность почвы, высота снежного покрова, относительная влажность, атмосферное давление.

Технология автоматической интеграции данных

Под каждый тип поступающей статистической информации в зависимости от её структуры разрабатывается модуль первичного разбора входных данных. После контроля целостности данных происходит их занесение в базу данных. Нужно отметить, что достаточно сложным процессом становится согласование справочников единиц административного деления между информационной системой источника данных и информационной системой интегратора. Это связано со множеством факторов, включая:

- наличие различных наборов уникальных идентификаторов единиц административного деления;
- изменение административно-территориального деления страны с течением времени (в том числе слияние субъектов и районов, изменения статуса);
- вариации в написании названий субъектов на русском языке;
- целесообразность одновременной обработки некоторых видов данных для двух и более единиц административного деления (например, Москвы и Московской обл.)

Таким образом, каждый источник данных требует разработки отдельной схемы согласования привязки к базе данных информационной системы интегратора. Последующее усвоение аналогичных данных производится в автоматическом режиме.

Анализ разнородных интегрированных данных требует унифицированной процедуры доступа к ним на основе параметризируемого интерфейса программирования приложений (англ. Application programming interface — API). В зависимости от типа данных реализация API может принимать различные формы. Статистические данные, имеющие привязку к единицам административного деления, непосредственно запрашиваются из таблиц базы данных по уникальному идентификатору. Метеорологические данные, имеющие локализацию в узлах регулярной сетки, могут быть приведены к привязке к единицам административного деления в момент запроса путём обработки значений в тех узлах, которые пространственно принадлежат запрашиваемой единице. Продукты обработки спутниковых данных не могут быть интерпретированы в момент запроса в силу большого числа пикселей изображения, принадлежащих запрашиваемой единице, поэтому требуется предварительная обработка таких данных.

Вычисление показателей на основе спутниковых данных (таких как вегетационные индексы, их отклонения от средних многолетних величин) производится заблаговременно для

минимизации времени на выполнение запроса при обращении к API. При этом в силу постоянного пополнения архива спутниковых данных в ЦКП «ИКИ-Мониторинг» необходимо, чтобы показатели, имеющие привязку к единицам административного деления и ко времени, также обновлялись. Такая обработка требует вычислительных ресурсов и производится в фоновом режиме с использованием системы управления очередью на исполнение. Результат предварительной обработки заносится в базу данных.

При пространственной интеграции данных на основе единиц административного деления существует возможность учёта неоднородности природных условий на этой территории, в частности наличия различных типов растительности. Таким образом, при получении оценки значений вегетационного индекса на территории административной единицы можно получить:

- среднее значение NDVI (Normalized Difference Vegetation Index — нормализованный разностный вегетационный индекс) на всей территории субъекта или района;
- отдельные средние значения NDVI на территориях, покрытых лесом, лугами, сельскохозяйственными угодьями и другими типами растительности.

Наличие множества измерений по спутниковым данным и по метеоданным за период времени позволяет при интеграции данных в заданные временные единицы получать не только средние значения, но и ряд других показателей, таких как накопленные величины, максимумы и минимумы, стандартные отклонения, отклонения от среднего многолетнего. Всё это позволяет привести к единообразному виду большое количество пригодных для анализа величин.

Использование унифицированного API для доступа к данным различной природы позволяет проводить их сравнительный анализ путём построения карт, графиков, экспорта табличных данных в единообразном интегрированном виде для последующего поиска корреляций.

Возможности визуализации и интерактивного анализа

Визуализация данных в привязке к административно-территориальным единицам возможна с помощью карт и графиков.

Рис. 1. Картограмма значений NDVI в административных районах

Картографический интерфейс позволяет отображать различные комбинации картографических слоёв, спутниковые изображения и результаты их обработки, метеорологические данные. Основным способом изображения статистических данных являются картограммы. Способ картограмм также применим и к любым другим характеристикам, вычисляемым на основе спутниковых и метеорологических данных. Таким образом, в разрезе административно-территориальных единиц могут быть показаны в том числе вегетационные индексы и климатические показатели (рис. 1 (см. с. 108), 2).

Рис. 2. Картограмма заболеваемости клещевым энцефалитом в субъектах РФ

Рис. 3. Плотность лесной растительности в ячейках регулярной сетки

Анализ данных, интегрированных в пространственные и временные единицы, возможен также с помощью ВІ-технологий (англ. Business intelligence) (Кобец и др., 2015). Интеграция

происходит в пространственных единицах различного уровня: федеральных округах, субъектах Российской Федерации, а также ячейках регулярной сетки. На *рис. 3* (см. с. 109) представлена интерактивная карта плотности растительного покрова территории РФ в ячейках регулярной сетки. Инструменты интерактивного анализа позволяют управлять принципами агрегации данных, включая периоды времени и интересующие классы растительности. Возможно представление данных в виде графиков и экспорт в табличном виде.

Примеры сравнения результатов обработки спутниковых данных и статистической информации

Анализ спутниковых данных проводился одновременно с анализом многолетних данных об урожайности сельскохозяйственных культур (Денисов и др., 2020). Соотношение значений NDVI, полученных в текущем году, со средними многолетними значениями индекса характеризует степень благоприятности нынешних условий для развития сельскохозяйственной растительности и позволяет судить об ожидаемой урожайности (*рис. 4*).

Рис. 4. Отклонение значений NDVI от средних многолетних значений

Сопоставление полученных данных о значениях NDVI с урожайностью озимых на уровне административных районов (по данным Единой межведомственной информационно-статистической системы — ЕМИСС, <https://fedstat.ru>) показало наличие зависимости между этими величинами (*рис. 5*, см. с. 111).

Технология автоматизированного картографирования типов земного покрова с использованием прибора MODIS (*англ.* Moderate Resolution Imaging Spectroradiometer) позволила получить серию карт растительности России (Уваров, Барталев, 2010). Оценка точности полученных карт проводилась путём оценки доли площади, занятой лесом в субъектах РФ, и сопоставления этих величин с данными Росстата о площади лесов по субъектам (*рис. 6*, см. с. 111).

Рис. 5. Связь значений NDVI озимых культур и урожайности озимой пшеницы

Рис. 6. Площади лесов в субъектах РФ по данным Росстата (PC) и карты TerraNorte RLC (RLC) (а). Распределение площади лесов в субъектах РФ (S) по относительной величине ошибки (б). В подписях точек указано количество субъектов

Заключение

Представленные в настоящей работе подходы обеспечивают совместный анализ статистических, спутниковых и метеорологических данных. Реализовано автоматическое получение специализированных продуктов данных на основе обработки спутниковых и метеоданных.

Специализированный блок инструментов в системе «Вега-Science» реализует следующую функциональность:

- возможности с работы со статистическими данными;
- возможности автоматической интеграции различных данных в разрезе территориальных единиц;

- возможности совместной визуализации и интерактивного анализа статистической информации и данных о состоянии окружающей среды, полученных на основе спутниковых и метеоданных.

Система функционирует с использованием ресурсов ЦКП «ИКИ-Мониторинг».

Работа выполнена при поддержке гранта Российского научного фонда № 19-75-20088 «Создание опирающейся на данные дистанционного зондирования Земли методологии анализа и прогнозирования влияния климатических и экологических факторов на заболеваемость природно-очаговыми инфекциями».

Литература

1. Денисов П. В., Середа И. И., Трошко К. А., Лупян Е. А., Толпин В. А., Плотников Д. Е. Особенности развития озимых и яровых культур на европейской территории России в сезоне 2019–2020 гг. // Современные проблемы дистанционного зондирования Земли из космоса. 2020. Т. 17. № 4. С. 306–311.
2. Кобец Д. А., Балашов И. В., Данилов И. Д., Лупян Е. А., Сычугов И. Г., Толпин В. А. Использование VI-технологий для создания инструментов для анализа данных спутникового мониторинга // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 4. С. 17–27.
3. Лупян Е. А., Прошин А. А., Бурцев М. А., Кашицкий А. В., Балашов И. В., Барталев С. А., Константинова А. М., Кобец Д. А., Мазуров А. А., Марченков В. В., Матвеев А. М., Радченко М. В., Сычугов И. Г., Толпин В. А., Уваров И. А. Опыт эксплуатации и развития центра коллективного пользования системами архивации, обработки и анализа спутниковых данных (ЦКП «ИКИ-Мониторинг») // Современные проблемы дистанционного зондирования Земли из космоса. 2019. Т. 16. № 3. С. 151–170. DOI: 10.21046/2070-7401-2019-16-3-151-170.
4. Платонов А. Е., Толпин В. А., Дубянский В. М., Уваров И. А., Титков А. В., Колясникова Н. М., Платонова О. В., Лупян Е. А., Малеев В. В. Связь климатических и экологических условий с заболеваемостью природно-очаговыми инфекциями: взгляд из космоса // Актуальные вопросы изучения особо опасных и природно-очаговых болезней: сб. статей конф. Новосибирск: Типография Прогресс, 2019. С. 227–236.
5. Прислегина Д. А., Дубянский В. М., Платонов А. Е., Малецкая О. В. Влияние природно-климатических факторов на эпидемиологическую ситуацию по природно-очаговым инфекциям: обзор // Инфекция и иммунитет. 2021. (В печати.)
6. Уваров И. А., Барталев С. А. Алгоритм и программный комплекс распознавания типов земного покрова на основе локально-адаптивной обучаемой классификации спутниковых изображений // Современные проблемы дистанционного зондирования Земли из космоса. 2010. Т. 7. № 1. С. 353–365.
7. Уваров И. А., Халикова О. А., Балашов И. В., Бурцев М. А., Лупян Е. А., Матвеев А. М., Платонов А. Е., Прошин А. А., Толпин В. А., Крашенинникова Ю. С. Организация работы с метеорологической информацией в информационных системах дистанционного мониторинга // Современные проблемы дистанционного зондирования Земли из космоса. 2013. Т. 10. № 2. С. 30–45.
8. Platonov A. E., Dubyansky V. M., Titkov A. V., Uvarov I. A., Tolpin V. A., Kolyasnikova N. M., Prislegina D. A., Raichich S. R., Maleev V. V. Remote sensing of innate killers // Intern. Symp. Tick-Borne Pathogens and Disease: Book of abstr. 8–11 сент. 2019, Vienna, Austria. Vienna: Robidruck Publ., 2019. P. 8.

Integration of satellite and meteorological data time series with statistical data in monitoring information systems

I. A. Uvarov¹, A. E. Platonov², A. V. Titkov², V. A. Tolpin¹, V. V. Maleev²

¹ Space Research Institute RAS, Moscow 117997, Russia

E-mail: smis@smis.iki.rssi.ru

² Central Research Institute of Epidemiology of the Federal Service on Customers' Rights Protection and Human Well-being Surveillance, Moscow 111123, Russia

E-mail: platonov@pcr.ru

The joint analysis of statistical data linked to administrative divisions and products of satellite data processing, as well as meteorological data, is of interest for identifying the relationships between the observed values. Integration of statistical data into information systems requires the development of special tools for automated processing, archiving, and the development of interactive analysis tools. They provide a harmonization of all types of source data to a comparable form, aggregation of a set of spatially distributed and time-repeated measurements to a number of integral characteristics. The paper presents the possibilities of solving this problem in the framework of the information system Vega-Science (a unique scientific installation, which is part of the IKI-Monitoring center for collective use). Examples of using the created solutions for analyzing data on the state of agricultural and forest vegetation, as well as on the spread of dangerous infectious diseases are given.

Keywords: remote monitoring, state of vegetation cover, natural focal infections, information system, meteorological data

Accepted: 08.12.2020

DOI: 10.21046/2070-7401-2020-17-7-105-114

References

1. Denisov P. V., Sereda I. I., Troshko K. A., Loupian E. A., Tolpin V. A., Plotnikov D. E., Osobennosti razvitiya ozimyykh i yarovyykh kul'tur na evropeiskoi territorii Rossii v sezone 2019–2020 gg. (Development of winter and spring crops in European Russia in the season 2019–2020), *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2020, Vol. 17, No. 4, pp. 306–311, DOI: 10.21046/2070-7401-2020-17-4-306-311.
2. Kobets D. A., Balashov I. V., Danilov I. D., Loupian E. A., Sychugov I. G., Tolpin V. A., Ispol'zovanie VI-tekhnologii dlya sozdaniya instrumentov dlya analiza dannykh sputnikovogo monitoringa (Using BI technologies to create analysis tools for satellite remote sensing data), *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2015, Vol. 12, No. 4, pp. 17–27.
3. Loupian E. A., Proshin A. A., Bourtsev M. A., Kashnitskiy A. V., Balashov I. V., Bartalev S. A., Konstantinova A. M., Kobets D. A., Mazurov A. A., Marchenkov V. V., Matveev A. M., Radchenko M. V., Sychugov I. G., Tolpin V. A., Uvarov I. A., Opyt ekspluatatsii i razvitiya tsentra kollektivnogo pol'zovaniya sistemami arkhivatsii, obrabotki i analiza sputnikovyykh dannykh (TsKP "IKI-Monitoring") (Experience of development and operation of the IKI-Monitoring center for collective use of systems for archiving, processing and analyzing satellite data), *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2019, Vol. 16, No. 3, pp. 151–170, DOI: 10.21046/2070-7401-2019-16-3-151-170.
4. Platonov A. E., Tolpin V. A., Dubyansky V. M., Uvarov I. A., Titkov A. V., Kolyasnikova N. M., Platonova O. V., Loupian E. A., Maleev V. V., Svyaz' klimaticheskikh i ekologicheskikh uslovii s zabolevaemost'yu prirodno-ochagovymi infektsiyami: vzglyad iz kosmosa (The relationship of climatic and ecological conditions with the incidence of natural focal infections: a view from space), *Aktual'nye voprosy izucheniya osobopasnykh i prirodno-ochagovykh boleznei* (Current problems in studying especially dangerous and natural focal diseases), Proc. Conf., Novosibirsk: Tipografiya Prodvizhenie, 2019, pp. 227–236.
5. Prislegina D. A., Dubyansky V. M., Platonov A. E., Maletskaya O. V., Vliyanie prirodno-klimaticheskikh faktorov na epidemiologicheskuyu situatsiyu po prirodno-ochagovym infektsiyam: obzor (Influence of natural and climatic factors on the epidemiological situation of natural focal infections: review), *Infektsiya i immunitet*, 2021. (In print)
6. Uvarov I. A., Bartalev S. A., Algoritm i programmnyi kompleks raspoznavaniya tipov zemnogo pokrova na osnove lokal'no-adaptivnoi obuchaemoi klassifikatsii sputnikovyykh izobrazhenii (The Algorithm and

- Software Suite for Land Cover Types Recognition Based on Locally-adaptive Supervised Classification of Satellite Imagery), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2010, Vol. 7, No. 1, pp. 353–365.
7. Uvarov I. A., Khalikova O. A., Balashov I. V., Bourtsev M. A., Loupian E. A., Matveev A. M., Platonov A. E., Proshin A. A., Tolpin V. A., Krashenninnikova Yu. S., Organizatsiya raboty s meteorologicheskoi informatsiei v informatsionnykh sistemakh distantsionnogo monitoringa (Meteorological data management in framework of the satellite monitoring information systems), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2013, Vol. 10, No. 2, pp. 30–45.
 8. Platonov A. E., Dubyansky V. M., Titkov A. V., Uvarov I. A., Tolpin V. A., Kolyasnikova N. M., Prislegina D. A., Raichich S. R., Maleev V. V., Remote sensing of innate killers, *Intern. Symp. Tick-Borne Pathogens and Disease*, Book of abstr., 8–11 Sept. 2019, Vienna, Austria, Vienna: Robidruck Publ., 2019, p. 8.