УДК 519.21

Ю.М. ЕРМОЛЬЕВ, Т.Ю. ЕРМОЛЬЕВА, П. ХАВЛИК, А. МОНЬЕ, Д. ЛЕКЛЕР, С. ФРИТЦ, М. ОБЕРШТАЙНЭР, С.В. КИРИЗЮК, Е.Н. БОРОДИНА

МЕТОДЫ РОБАСТНОГО РАЗУКРУПНЕНИЯ ДАННЫХ И ПРОЕКЦИЙ ПРИ НЕОПРЕДЕЛЕННОСТЯХ: ИССЛЕДОВАНИЯ ИЗМЕНЕНИЙ ЗЕМЕЛЬНОГО ПОКРОВА И ЗЕМЛЕПОЛЬЗОВАНИЯ[1]

Аннотация. Рассмотрены взаимозависимости между системами землепользования на локальном, национальном и глобальном уровнях, которые обусловливают необходимость разработки новых методов системного анализа для интеграции моделей землепользования разных масштабов. Разработаны новые общие подходы получения разукрупненных оценок на основании принципа кросс-энтропии, которые робастны относительно множества возможных праеров. Методы робастного разукрупнения учитывают так называемые небайесовские неопределенности, т.е. неполноту, отсутствие, ошибки в данных. В многочисленных практических исследованиях, проведенных в Китае, странах Африки, Бразилии, Украине, предложенные подходы позволили получить локальные прогнозы развития и изменения землепользования, соответствующие реальным тенденциям и ожиданиям.

Ключевые слова: планирование устойчивого развития, неопределенности, робастное разукрупнение, локально-глобальные зависимости.

ВВЕДЕНИЕ

Изменение землепользования и земного покрова — ключевые факторы глобальных и локальных изменений окружающей среды [1]. На протяжении последних десятилетий было разработано множество моделей для исследования изменений землепользования и земного покрова [2–4] в целях анализа возможных сценариев землепользования на глобальном и региональных уровнях. Некоторые из этих моделей уделяют внимание только определенным локальным аспектам землепользования, например развитию лесных хозяйств и использованию пахотных земель, расширению городских территорий. Локальные модели и исследования такого рода не учитывают взаимодействия глобальных и локальных факторов и тенденции. И, наоборот, агрегированные глобальные модели землепользования часто не в состоянии учесть локальные тревожные показатели, например ограниченность водных ресурсов, что может привести к несбалансированному планированию и избыточному потреблению ресурсов на местах. Локальные изменения, как правило, являются неизбежным следствием региональных, национальных и глобальных тенденций, сельскохозяйственных реформ, политических договоров и соглашений, стимулирующих спрос на продовольственные продукты, производство сырья для биотоплив, что, в свою очередь, обусловлено взаимодействием между социальными, экономическими, демографическими, климатическими факторами и процессами. Взаимозависимости между системами землепользования на локальном, национальном и глобальном уровнях обусловливают необходимость разработки новых методов системного анализа для интеграции, агрегирования и разукрупнения результа-

[1] Исследования проводятся в рамках EU-проектов ECONADAPT (603906), TRANSMANGO (613532), AGRICISTRADE (612755) EU FP7, а также научного проекта по разработке новаторских методологий и приложений, исследующих робастные решения для долгосрочного согласованного планирования безопасного снабжения продовольствием, энергией, водой, проводимого совместно Международным институтом прикладного системного анализа (Лаксенбург, Австрия) и Национальной академией наук Украины.

© Ю.М. Ермольев, Т.Ю. Ермольева, П. Хавлик, А. Монье, Д. Леклер, С. Фритц, М. Оберштайнэр, С.В. Киризюк, Е.Н. Бородина, 2017

тов моделей землепользования разных масштабов. В настоящей статье предлагаем методы робастного разукрупнения, которые будут применены в следующей публикации в целях проведения исследований и получения прогнозов на требуемом пространственном разрешении.

В разд. 1 приводятся обоснования для разработки модели пространственного оценивания и разукрупнения данных, основанной на принципе кросс-энтропии. Необходимость комплексного пространственно-детализированного анализа последствий глобальных изменений порождает ряд новых проблем, требующих разработки нетрадиционных процедур оценивания. В то время как методы статистического оценивания предполагают возможность получения повторяющихся наблюдений, для задачи разукрупнения доступны лишь агрегированные данные и достаточно ограниченные (часто с ошибками) локальные наблюдения. В этой связи цель задачи разукрупнения состоит в получении правдоподобных оценок на локальном уровне исходя из глобальных тенденций и проекций агрегированных моделей, используя все доступные данные, наблюдаемые и ненаблюдаемые переменные, зависимости между переменными, заключения экспертов, а также результаты, полученные в других моделях. В разд. 1 также сформулирована общая задача дезагрегации, и на стилизованном примере проиллюстрировано применение принципа кросс-энтропии для разукрупнения данных по производству сельскохозяйственной продукции. Принцип кросс-энтропии предполагает некоторое начальное (априорное распределение) производство на местах, т.е. праер.

В разд. 2 приведен метод последовательной перебалансировки (разукрупнения) для оценивания ненаблюдаемых значений, потоков, маргинальных изменений в системах, предложенный ленинградским архитектором Желиховским в 1930 г. еще до появления теории информации Шеннона [5]. Решение предложенного алгоритма сходится к решению соответствующей задачи кросс-энтропии. Алгоритм основан на поведенческих критериях выбора праера, т.е. начального распределения, которыми будем руководствоваться в дальнейшем. Важным обоснованием (см. разд. 3) для использования подхода кросс-энтропии является связь этого метода с одним из основных в статистике методом максимального правдоподобия, предложенным в 1922 г. Фишером. Решение задачи минимизации кросс-энтропии минимизирует также так называемое информационное расстояние Кульбака–Ляйблера [6] между наблюдаемым и неким желаемым распределением при дополнительных ограничениях, например на использование ресурсов, нормы загрязнения почвы и воды, эмиссии парниковых газов. Данная мера расстояния может интерпретироваться как прибыль (польза), полученная от замены наблюдаемого распределения на оцененное распределение, удовлетворяющее всем необходимым ограничениям. Практические исследования свидетельствуют, что априорное распределение может зависеть от различных случайных параметров, таких как, например, урожайность, цена, уровень эмиссий и т.д. Таким образом, вместо единственного праера, как это принято в традиционном подходе, может существовать допустимое множество праеров, и в этом случае задача разукрупнения состоит в том, чтобы получить локальные оценки, используя информацию, заключенную во всех праерах, т.е. робастные оценки. Возможные подходы к робастному разукрупнению приведены в разд. 4. В заключении статьи суммируются общие выводы, которые будут учтены в практических применениях предложенных методов.

1. ОБЩАЯ МОДЕЛЬ ИЗМЕНЕНИЯ РАЗРЕШЕНИЯ ЗЕМЛЕПОЛЬЗОВАНИЯ

Рассмотрим фрагмент общей модели разукрупнения, детально описанной в [7]. Проиллюстрируем ее на примере разукрупнения агрегированных данных о пахотных землях. Используемая в модели информация может быть обобщена следующим образом. Оценка площади пахотной земли a_j в пространственной точке j, $j = \overline{1, n}$, может быть получена с помощью карт земного покрова, из статистических сборников или как результаты глобальных моделей (GLOBIOM [8]). Оценка пригодности земли характеризует потенциально достижимый уровень продуктивности y_{ij} определенного типа землепользования i, $i = \overline{1, m}$, например пахотных земель, пастбищ, лесных хозяйств и т.д., в разных пространственных точках j. Географически детализированная информация о биофизической пригодности и потенциальной продуктивности земли в разрезе различных сельскохозяйственных культур может быть получена на основании исторических данных, а также с помощью биофизических моделей таких, например, как модели EPIC [9] или GAEZ [10].

Цель задачи разукрупнения состоит в определении такого уровня x_{ij} землепользования i в точке j, которое удовлетворяло бы основным балансовым уравнениям:

$$\sum_{i=1}^{m} x_{ij} = a_j, \; j = \overline{1, n}, \tag{1}$$

$$\sum_{j=1}^{m} y_{ij} x_{ij} \le d_i, \; i = \overline{1, m}, \tag{2}$$

где переменная x_{ij} и параметры a_j, y_{ij}, d_i имеют положительные значения. Уравнение (1) устанавливает ограничение на совокупное землепользование в точке j, тогда как уравнение (2) накладывает ограничение на количество произведенной продукции d_i при землепользовании i в соответствии с производительностью y_{ij}. Данные или проекции спроса d_i, например на производство зерновых или древесины, доступны из официальной статистики или могут быть подсчитаны с помощью агрегированных моделей, подобных GLOBIOM [8].

Можно ввести новые переменные z_{ij}, характеризующие распределение площадей ($0 \le z_{ij} \le 1$) под определенными типами землепользования i в точке j. Таким образом, $x_{ij} = a_j z_{ij}$, и ограничения (1), (2) могут быть записаны как

$$\sum_{i=1}^{m} z_{ij} = 1, \; j = \overline{1, n}, \tag{3}$$

$$\sum_{j=1}^{n} a_{ij} z_{ij} \le d_i, \; i = \overline{1, m}, \; a_{ij} = a_i y_{ij}. \tag{4}$$

Эта модификация ограничений (1), (2) позволяет использовать принцип кросс-энтропии (и вероятностные доводы) для определения оптимального решения z_{ij}. В общем случае может существовать множество решений z_{ij}, $i = \overline{1, m}$, $j = \overline{1, n}$, удовлетворяющих уравнениям (3), (4), например когда система (3), (4) недоопределена. В этом случае необходимо из множества решений выбрать одно исходя из дополнительного критерия. Ключевым аспектом в выборе такого критерия является то, что уже имеется так называемое априорное (желаемое) распределение z_{ij}, $i = \overline{1, m}$, $j = \overline{1, n}$, которое вызывает минимальные негативные (или максимальные позитивные), в некотором смысле, последствия. Например, интуиция подсказывает, что априорное распределение может зависеть от рентабельности, чистой выручки или максимальной чистой приведенной стоимости за активность в каждой точке пространства. Априорное распределение можно рассчитать, используя доступные данные (карты) землепользования, карты климатических и биофизических условий в пространственных точках либо исходя из условий формирования спроса или систем хозяйствования. Например, априорное распределение $q_{ij}=\dfrac{\alpha_{ij}P_i y_{ij}}{\sum_i \alpha_{ij}P_i y_{ij}}$, $\sum_j q_{ij}=1$, может служить начальной оценкой площади землепользования i в пространственной единице j, где P_i — цена продукции i-го типа землепользования, y_{ij} — урожайность i-го типа землепользования в пространственной единице j. В условиях отсутствия дополнительной информации априорным может быть наименее информативное равномерное распределение, при этом проблема разукрупнения по-прежнему включает ограничения на всех уровнях и из всех источников. Таким образом, задача разукрупнения будет корректировать априорное распределение в целях выполнения всех имеющихся ограничений, в данном случае (3), (4). В разд. 2 обсуждаются возможные подходы для выбора априорных распределений. Если априорное распределение $q_{ij}>0$, $i=\overline{1,m}$, $j=\overline{1,n}$, известно, то оценки z_{ij} кросс-энтропии определяются посредством минимизации функции кросс-энтропии

$$\sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij}\ln\frac{z_{ij}}{q_{ij}} \tag{5}$$

с учетом (3), (4). Функция (5) определяет расстояние Кульбака–Ляйблера [6] между распределениями z_{ij} и q_{ij}, т.е. распределение z_{ij}, полученное в результате решения задачи минимизации кросс-энтропии, будет ближайшим к априорному распределению q_{ij}, при этом выполняются ограничения (3), (4).

2. ПОВЕДЕНЧЕСКИЙ АЛГОРИТМ ПОСЛЕДОВАТЕЛЬНОГО РАЗУКРУПНЕНИЯ ДАННЫХ. СХОДИМОСТЬ РЕШЕНИЯ К РЕШЕНИЮ СООТВЕТСТВУЮЩЕЙ ЗАДАЧИ МИНИМИЗАЦИИ КРОСС-ЭНТРОПИИ

Рассмотрим пример оценивания потоков в сетях. Интересно отметить, что в 1930 г., т.е. еще до появления теории информации, предложенной Шенноном [5], ленинградский архитектор Желиховский разработал алгоритм последовательной перебалансировки (разукрупнения) для оценивания ненаблюдаемых значений и потоков в сетях. Решение предложенного алгоритма сходится к решению соответствующей задачи кросс-энтропии. Алгоритм был предложен для изучения и планирования пассажиропотоков между районами крупного города. Метод применим также для исследования возможных потоков мигрантов, транспортных потоков, а также для изучения возможных изменений типов землепользования. Подход, предложенный Желиховским, оценивает возможные потоки в соответствии с аггрегированными данными о выходах (отбытиях) a_i из пункта i и приходах (прибытиях) b_j в пункт j. Выходные и входные потоки можно интерпретировать как маргинальные изменения на местах (например, изменение количества мигрантов, объема поставок, земли в определенном землепользовании и т.д.). Таким образом, неизвестные потоки x_{ij} удовлетворяют уравнению $\sum_{j=1}^{n} x_{ij}=a_i$, $i=\overline{1,m}$, $\sum_{i=1}^{m} x_{ij}=b_j$, $j=\overline{1,n}$, т.е. в этом случае имеем огранничения типа (1), (2) при $y_{ij}=1$, $i=\overline{1,m}$, $j=\overline{1,n}$.

Предположим, что имеем априорно заданные вероятностности q_{ij} того, что поток из пункта i направится в пункт j. Например, в соответствии с поведенческими принципами вероятности q_{ij} могут определяться обратно пропорционально расстоянию r_{ij} от пункта i до пункта j, $q_{ij}=(r_{ij})^{-1}/(\sum_j r_{ij})^{-1}$. Рассмотрим следующую итеративную процедуру оценивания.

1. Если пассажир из пункта i выберет пункт назначения j с априорной вероятностью q_{ij}, $\sum_j q_{ij} = 1$, то ожидаемый поток из i в j определится как $x_{ij}^0 = a_i q_{ij}$. Очевидно, что баланс $\sum_j x_{ij}^0 = a_i$, $i = \overline{1, m}$, выходных потоков сохраняется. Для входных потоков может быть две ситуации: недооценка $\sum_i x_{ij}^0 > b_j$ или переоценка $\sum_i x_{ij}^0 < b_j$ суммарного потока b_j.

2. Подсчитаем относительный дисбаланс $\beta_j^0 = b_j / \sum_i x_{ij}^0$ и новое значение потока $z_{ij}^0 = x_{ij}^0 \beta_j^0$, $i = \overline{1, m}$, $j = \overline{1, n}$.

3. Тогда выполняется условие $\sum_i z_{ij}^0 = b_j$, $j = \overline{1, n}$, но оценка z_{ij}^0 может привести к переоцениванию или недооцениванию суммарного выходящего потока (отправлений) a_i из i. На этом шаге пересчитаем дисбаланс $\alpha_i^0 = a_i / \sum_j z_{ij}^0$ и найдем значения потоков $x_{ij}^1 = z_{ij}^0 \alpha_i^0$, которые удовлетворяют условию суммарного потока $\sum_j x_{ij}^1 < a_i$.

Доказано, что алгоритм Желиховского вычисляет последовательность потоков $x_{ij}^0, x_{ij}^1, \ldots, x_{ij}^k, \ldots$, которая сходится к решению задачи минимизации кросс-энтропии (5) при ограничениях (1), (2) для всех i и j. Важно отметить, что метод можно представить в виде стандартной статистической процедуры последовательного пересчета априорного распределения q_{ij} после получения (или наблюдения) дополнительной информации о системных дисбалансах β_j. Действительно, можно представить x_{ij}^1 как $x_{ij}^1 = a_i q_{ij}^1$ с праером $q_{ij}^1 = (q_{ij} \beta_j^0) / (\sum_j q_{ij} \beta_j^0)$, $i = \overline{1, m}$, $j = \overline{1, n}$. Предположим, что значение потока $x^k = \{x_{ij}^k\}$ подсчитано. Тогда на шаге k необходимо подсчитать $\beta_j^k = b_j / \sum_i x_{ij}^k$ и получить $x_{ij}^{k+1} = a_i q_{ij}^{k+1}$, $q_{ij}^{k+1} = (q_{ij} \beta_j^k / \sum_j q_{ij} \beta_j^k)$, $i = \overline{1, m}$, $j = \overline{1, n}$, и т.д. В таком виде представленную процедуру можно рассматривать как процедуру распределения пассажиров (мигрантов, производства, спроса) a_i из пункта $i = \overline{1, m}$ между пунктами $j = \overline{1, n}$, применяя байесовское правило пересчета априрного распределения q_{ij}: $q_{ij}^{k+1} = q_{ij} \beta_j^k / \sum_j q_{ij} \beta_j^k$, $q_{ij}^0 = q_{ij}$. Доказательство сходимости данного метода было приведено Брегманом [11] с использованием сложных и длительных доказательств на основании отображений и проекций, возникающих в связи с транспортными ограничениями. Для более общих ограничений аналогичная процедура на основе принципа двойственности была разработана, доказана и применена [12] для исследования устойчивого развития (расширения и размещения) сельскохозяйственного производства в Китае и Украине.

3. СВЯЗЬ ЗАДАЧИ МИНИМИЗАЦИИ КРОСС-ЭНТРОПИИ С МЕТОДОМ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Рассмотрим связь принципа кросс-энтропии и ключевого в статистике принципа максимального правдоподобия. Имеем задачу кросс-энтропии для распределения $(p_1, \ldots, p_r)$ с праером $(q_1, \ldots, q_r)$

$$\sum_{j=1}^{r} p_j \ln \frac{p_j}{q_j} = \sum_{j=1}^{r} p_j \ln p_j - \sum_{j=1}^{r} p_j \ln q_j .$$

Покажем, что минимизация $\sum_{j=1}^{r} p_j \ln p_j$ эквивалентна решению задачи maximin функции правдоподобия. Рассмотрим ситуацию, аналогичную задаче, сформулированной в разд. 1. Предположим, что в основе наблюдаемых величин $\xi^1, \ldots, \xi^N$ лежит реальная случайная величина ξ, и неизвестное истинное распределение вероятности этой величины характеризуется величинами $p_1^*, \ldots, p_r^*$ такими, что $\mathrm{Prob}[\xi = \xi^j] = p_j^*$.

В статистическом оценивании имеющаяся информация характеризуется случайной выборкой $\xi^1, \ldots, \xi^N$ из N независимых наблюдений величины ξ с вероятностями $(p_1^*, \ldots, p_r^*)$. Оценку максимального правдоподобия вероятностей $(p_1^*, \ldots, p_r^*)$ в этом случае получаем в результате максимизации функции правдоподобия (т.е. максимизации вероятности наблюдения выборки $\xi^1, \ldots, \xi^N$)

$$\prod_{k=1}^{N} \mathrm{Prob}[\xi = \xi^k] = \prod_{j=1}^{r} p_j^{\upsilon_j} \tag{6}$$

при ограничениях $\sum_{j=1}^{r} p_j = 1$, $p_j > 0$, $j = \overline{1, r}$, где υ_j определяет частоту появления величины ξ_j в выборке, $\sum_{j=1}^{r} \upsilon_j = N$. Функция $\ln y$ является монотонно возрастающей по y, поэтому максимизация (6) эквивалентна максимизации логарифма функции максимального правдоподобия $\ln \prod_{j=1}^{r} p_j^{\upsilon_j} = \sum_{j=1}^{r} \upsilon_j \ln p_j$ или функции эмпирического среднего, т.е. нормированной числом наблюдений N, $\sum_{j=1}^{r} \upsilon_j = N$,

$$\frac{1}{N} \sum_{j=1}^{r} \upsilon_j \ln p_j . \tag{7}$$

В результате получим непрерывную, строго выпуклую функцию кросс-энтропии на множестве R^n, определяемом линейными ограничениями. Используя функцию Лагранжа, можно получить известный результат, согласно которому решение (единственное), максимизирующее (7), является эмпирическим распределением вероятности

$$p_j^N = \upsilon_j / N, \quad j = \overline{1, r}. \tag{8}$$

Рассмотрим задачу с иной точки зрения. Функция (7) представляет эмпирическое среднее математического ожидания

$$E \ln p_\xi = \sum_{j=1}^{r} p_j^* \ln p_j , \tag{9}$$

где неизвестное распределение вероятности p_j^* аппроксимируется частотой υ_j / N, подсчитанной на основании имеющихся наблюдений выборки $\xi^1, \ldots, \xi^N$. При решении задачи разукрупнения информация о неизвестном истинном распределении вероятности p_j^*, $j = \overline{1, r}$, задана не выборкой наблюдений, а с помощью таких ограничений, как (3) и (4), т.е. в общем случае

$p^* \in P$, где P представляет множество всех возможных распределений. Если $u = (u_1, \ldots, u_r) \in P$, то $\sum_{j=1}^{r} u_j \ln p_j$ является аппроксимацией функции математического ожидания (9), аналогичной эмпирической функции (7). Оценка maximin на множестве P приводит к максимизации функции $\min_{u \in P} \sum_{j=1}^{r} u_j \ln p_j$.

Из условия строгой выпуклости этой фукции следует, что проблема эквивалентна минимизации функции $U(u) = \max_{p \in P} \sum_{j=1}^{r} u_j \ln p_j$ при $u \in P$. Отметим, что $\min U(u)$ при $u \in P$ можно записать как

$$\min_{u \in P} \max_{p \in P} \sum_{j=1}^{r} u_j \ln p_j = \min_{u \in P} \sum_{j=1}^{r} u_j \ln u_j$$

в силу неравенства

$$\sum\nolimits_{j=1}^{r} u_j \ln p_j - \sum\nolimits_{j=1}^{r} u_j \ln u_j = \sum\nolimits_{j=1}^{r} u_j \ln \frac{p_j}{u_j} < \ln \sum\nolimits_{j=1}^{r} u_j \frac{p_j}{u_j} = 0.$$

Другими словами, оценка максимального правдоподобия действительно минимизирует $\sum\nolimits_{j=1}^{r} u_j \ln u_j$ или $-\sum\nolimits_{j=1}^{r} u_j \ln u_j$. Для заданного праера $q = (q_1, \ldots, q_r)$ это утверждение также справедливо и в случае минимизации кросс-энтропии $\sum\nolimits_{j=1}^{r} u_j \ln \frac{u_j}{q_j} = \sum\nolimits_{j=1}^{r} u_j \ln u_j - \sum\nolimits_{j=1}^{r} u_j \ln q_j$, представляющей сумму $\sum\nolimits_{j=1}^{r} u_j \ln u_j$ и $-\sum\nolimits_{j=1}^{r} u_j \ln q_j$, предполагающей, что большему значению праера q_j будет соответствовать большее значение u_j.

4. РОБАСТНОЕ РАЗУКРУПНЕНИЕ ПРИ НЕОПРЕДЕЛЕННОСТЯХ В ПРАЕРАХ

Рассмотрим задачу разукрупнения (3)–(5): минимизировать функцию кросс-энтропии $\sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1}^{n} z_{ij} \ln \frac{z_{ij}}{q_{ij}}$ при ограничениях $\sum\nolimits_{i=1}^{m} z_{ij} = 1$, $j = \overline{1, n}$, $\sum\nolimits_{j=1}^{n} \alpha_{ij} z_{ij} \le d_i$, $i = \overline{1, m}$, где значения z_{ij} положительны, и допустимое множество решений z_{ij} обозначим Z.

В традиционной задаче (3)–(5) задан единственный праер q_{ij}. Однако практические исследования свидетельствуют, что априорное распределение q_{ij} может зависеть от различных случайных параметров и, таким образом, может быть неоднозначно определено. Поэтому в общем случае вместо единственного праера q_{ij} можно предполагать, что существует множество праеров Q, $q \in Q$, $q = \{q_{ij}, i = \overline{1, m}, j = \overline{1, n}\}$. Основная задача состоит в том, чтобы найти решение задачи (3)–(5), в некотором смысле робастное относительно всех праеров q из Q. Рассмотрим два важных случая.

1. Сложные априорные распределения. Предположим, что существует конечное множество Q возможных праеров $q^s = \{q_{ij}^s, i = \overline{1, m}, j = \overline{1, n}, s = \overline{1{:}S}\}$, которые определяют множество функций кросс-энтропии $F^s(z) = \sum_i \sum_j z_{ij} \ln \frac{z_{ij}}{q_{ij}^s}$.

Сценарии праеров $s = \overline{1, S}$ могут соответствовать различным наблюдениям или альтернативным экспертным оценкам параметров α_{ij}, P_i, y_{ij}, определяющим априорное распределение, как это обсуждается в разд. 1. Достоверность сценариев $s = \overline{1, S}$ можно задавать с помощью вероятностей (весов) $\gamma_s > 0$, $\gamma_s \le 1$, $\sum_s \gamma_s = 1$, отражающих степень уверенности в том, что s это истинный сценарий. В случае полной неопределенности можно предположить, что все сценарии одинаково вероятны, т.е. $\gamma_s = 1/S$, $s = \overline{1, S}$. С использованием весов γ_s задача переформулируется в стандартном виде (3)–(5) при $q_{ij} = \sum_{s=1}^{S} \gamma_s q_{ij}^s$.

2. Робастные небайесовские априорные распределения. Кроме известной аггрегированной информации, заданной допустимыми ограничениями (3), (4), априорное распределение $q = \{q_{ij}, i = \overline{1, m}, j = \overline{1, n}\}$ использует важную начальную гипотезу о распределении z_{ij}, удовлетворяющем этим ограничениям. Для конечного множества Q априорных распределений q^s, $s = \overline{1, S}$, функция

$$F(z) = \max_s \sum_{i, j} z_{ij} \ln (z_{ij} / q_{ij}^s) = \max_s \sum_{ij} z_{ij} \ln (z_{ij} / q_{ij}^s(z))$$

определяет максимальное расстояние (в смысле информационного расстояния Кульбака–Ляйблера [6] до множества распределений Q от искомого распределения $z = \{z_{ij}, i = \overline{1, m}, j = \overline{1, n}\}$, удовлетворяющего допустимым ограничениям, где $q(z)$ — оптимальное априорное распределение. Таким образом, минимизацию $F(z)$ можно рассматривать как небайесовскую минимизацию кросс-энтропии с праером $q(z)$, зависящим от z. Покажем, что минимизация $F(z)$ эквивалетна минимизации функции

$$F(z) = \max_{\gamma} \sum_s \gamma_s \sum_{i, j} z_{ij} \ln (z_{ij} / q_{ij}^s), \quad \sum_{s=1}^{S} \gamma_s = 1, \ \gamma_s \ge 0.$$

Эта формулировка позволяет рассмотреть более общую задачу робастного разукрупнения на основе принципа кросс-энтропии с использованием весов $\gamma = (\gamma_1, \ldots, \gamma_S)$, удовлетворяющих некоторым общим ограничениям типа $\gamma \in H$, отражающим наличие информации относительно γ. Так, в определенных случаях можно считать, что суммарный вес некоторого подмножества сценариев превышает суммарный вес другого подмножества, например $\gamma_s \ge \gamma_t$, $\gamma_s + \gamma_t \ge 1/2$ для заданных сценариев s и t. Тогда достаточно общая задача робастной минимизации смешанной задачи кросс-энтропии формулируется как минимизация функции

$$F(z) = \max_{\gamma \in H} \sum_{s=1}^{S} \gamma_s \sum_{i, j} z_{ij} \ln (z_{ij} / q_{ij}^s) =$$

$$= \sum_{ij} z_{ij} \ln z_{ij} - \min_{\gamma \in H} \sum_{s=1}^{S} \gamma_s (\sum_{ij} z_{ij} \ln q_{ij}^s), \tag{10}$$

где z принадлежит допустимому множеству Z, заданному ограничениями типа (3), (4). Функция (10) линейна по γ и строго выпукла по z, но, к сожалению, она не является выпуклой функцией по обеим переменным (z, γ). Тем не менее, задачу минимизации (10) можно переформулировать, как задачу выпуклой оптимизации, сформулировав задачу, двойственную к задаче минимизации по γ:

$$\min_{\gamma \in H} \sum_{s=1}^{S} \gamma_s (\sum_{ij} z_{ij} \ln q_{ij}^s). \tag{11}$$

Для упрощения обсуждения предположим, что множество H задано линейными ограничениями

$$\sum_{s=1}^{S} \beta_{ks} \gamma_s \geq \delta_k, \quad k = \overline{1, K}, \tag{12}$$

для определенных β_{ks}, δ_k. Если для некоторого фиксированного Z определим, что $c_s := \sum_{ij} z_{ij} \ln q_{ij}^s$, тогда минимизация задачи (11), (12) относительно γ может быть записана как минимизация некоторой линейной (при фиксированном z) функции

$$\min_{\gamma \in H} \sum_{s=1}^{S} \gamma_s c_s, \quad c_s = \sum_{ij} z_{ij} \ln q_{ij}^s \tag{13}$$

при ограничениях (12). Двойственная задача формулируется как задача максимизации функции

$$\max_{u} \sum_{k=1}^{K} \delta_k u_k \tag{14}$$

при ограничениях $\sum_{k=1}^{K} \beta_{ks} u_k \leq c_s$, где $u = (u_1, \ldots, u_k)$, $u_k \geq 0$, $k = \overline{1, K}$, представляют двойственные переменные. В соответствии с теорией двойственности задача минимизации (12), (13) эквивалентна задаче максимизации (14). Таким образом, задача минимизации функции (10) может быть сформулирована как задача минимизации функции

$$\sum_{ij} z_{ij} \ln z_{ij} - \max_{u} \sum_{k} \delta_k u_k$$

относительно (z, u) при ограничениях

$$\sum_{k} \beta_{ks} u_k \leq \sum_{ij} z_{ij} \ln q_{ij}^s, \quad s = \overline{1, S}, \quad z \in Z, \tag{15}$$

т.е. минимизации функции

$$\sum_{ij} z_{ij} \ln z_{ij} - \sum_{k} \delta_k u_k$$

относительно (z, u) при ограничениях (15).

ЗАКЛЮЧЕНИЕ

В статье предложены подходы для разукрупнения данных и проекций различных экономических, социальных, экологических и других показателей в ситуациях, когда данные и проекции могут характеризоваться неопределенностями. С учетом неопределенностей и возможных ошибок, а также неполноты данных предложенные методы позволяют получить локальные оценки, используя все источники информации, т.е. экспертные оценки, балансовые уравнения, данные разного характера и на разных уровнях разрешения. Таким образом, в отличие от байесовских подходов к оцениванию, когда информацию об оцениваемой величине можно получать непосредственно из последовательных наблюдений, предполагая некоторую реальную статистическую модель (т.е. модель регрессии), предложенные подходы оперируют нестандартными случаями, когда наблюдения могут отсутствовать, а статистическая модель неизвестна. В многочисленных практических исследованиях, проведенных в Китае, странах Африки, Бразилии, Украине, предложенные подходы позволили получить локальные прогнозы развития и изменения землепользования, соответствующие реальным тенденциям и ожиданиям.

СПИСОК ЛИТЕРАТУРЫ

1. Lambin E. F., Meyfroidt P. Global land use change, economic globalization, and the looming land scarcity. *Proc. Natl. Acad. Sci. USA*. 2011. Vol. 108, N 9. P. 3465–3472.

2. Agarwal C., Green G.M., Grove J.M., Evans T.P., Schweik C.M. A review and assessment of land-use change models: Dynamics of space, time, and human choice. *Gen. Tech. Rep. NE-297. Newton Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station*. 2002. 61 p.
3. Chomitz K. M., Gray D. A. Roads, land use, and deforestation: A spatial model applied to Belize. *World Bank Econ. Rev*. 1996. Vol. 10. P. 487–512.
4. Lotze-Campen H., Popp A., Beringer T., Müller C., Bondeau A., Rost S., Lucht W. Scenarios of global bioenergy production: the trade-offs between agricultural expansion, intensification and trade. *Ecological Modeling*. 2010. Vol. 221. 2188–2196.
5. Shannon C.E. A mathematical theory of communication. *Bell System Technical Journal*. 1948. Vol. 27. P. 379–423, 623–659.
6. Kullback J. Information theory and statistics. New York: John Wiley and Sons, 1959. 395 p.
7. Zagorodny A.G., Ermoliev Y.M., Bogdanov V.L. Integrated management, security and robustness. *Publ. by Committee for Systems Analysis and Presidium of National Academy of Sciences, Ukraine — National Member Organization of the International Institute for Applied Systems Analysis* (IIASA). ISBN 978-966-02-7376-4. Kyiv. 2014.
8. Havlík P., Schneider U., Schmid E., Böttcher H., Fritz S., Skalsky R., Aoki K., De Cara S., Kindermann G., Kraxner F., Leduc S., McCallum I., Mosnier A., Sauer T., Obersteiner M. Global land-use implications of first and second generation biofuel targets. *Energy Policy*. 2011. Vol. 39. P. 5690–5702.
9. Izaurralde R.C., Williams J.R., McGill W.B., Rosenberg N.J., Jakas M.C.Q. Simulating soil C dynamics with EPIC: Model description and testing against long-term data. *Ecological Modeling*. 2006. Vol. 192. P. 362–384.
10. Fischer G., van Velthuizen H.T., Shah M.M., and Nachtergaele F.O. Global agro-ecological assessment for agriculture in the 21st century: Methodology and results. *Research Report RR-02-02. International Institute for Applied Systems Analysis*. Laxenburg, Austria. 2002.
11. Bregman L.M. Proof of the convergence of Sheleikhovskii's method for a problem with transportation constraints. *USSR computational mathematics and mathematical physic*s. 1967. Vol. 7, N 1. P. 191–204.
12. Fischer G., Winiwarter W., Ermolieva T., Cao G.-Y., Qui H., Klimont Z., Wiberg D. and Wagner F. Integrated modeling framework for assessment and mitigation of nitrogen pollution from agriculture: Concept and case study for China. *Agriculture, Ecosystems & Environment*. 2010. Vol. 136, N 1. P. 116–124.

Надійшла до редакції 19.07.2016

Ю.М. Єрмольєв, Т.Ю. Єрмольєва, П. Хавлик, А. Моньє, Д. Леклер, С. Фрітц, М. Оберштайнер, С.В. Киризюк, О.М. Бородіна

МЕТОДИ РОБАСТНОГО РОЗУКРУПНЕННЯ ДАНИХ І ПРОЕКЦІЙ ПРИ НЕВИЗНАЧЕНОСТЯХ: ДОСЛІДЖЕННЯ ЗМІН ЗЕМЕЛЬНОГО ПОКРИВУ ТА ЗЕМЛЕКОРИСТУВАННЯ

Анотація. Розглянуто взаємозалежності між системами землекористування на локальному, національному та глобальному рівнях, які обумовлюють необхідність розроблення нових методів системного аналізу для інтеграції моделей землекористування різних масштабів. Розроблено нові загальні підходи одержання розукрупнених оцінок на підставі принципу крос-ентропії, які є робастними відносно чисельності можливих праерів. Методи робастного розукрупнення враховують так звані небаєсівські невизначеності, тобто неповноту або відсутність даних, помилки у них. У чисельних практичних дослідженнях, проведених у Китаї, країнах Африки, Бразилії та Україні, запропоновані підходи дозволили отримати локальні прогнози розвитку і зміни землекористування відповідно до реальних тенденцій і очікувань.

Ключові слова: планування сталого розвитку, невизначеності, робастні розукрупнення, локально-глобальні залежності.

Y.M. Ermoliev, T.Y. Ermolieva, P. Havlík, A. Mosnier, D. Leclere, S. Fritz, M. Obersteiner, S.V. Kyryzyuk, O.M. Borodina

ROBUST DOWNSCALING APPROACHES TO DISAGGREGATION OF DATA AND PROJECTIONS UNDER UNCERTAINTIES: CASE OF LAND USE AND LAND USE CHANGE SYSTEMS

Abstract. The interdependencies among land use systems at national and global levels motivate the development of advanced systems analysis approaches for integration of land use models operating at different scales. The paper develops novel general approaches based on cross-entropy principle for downscaling aggregate data and projections, which are robust with respect to feasible priors. Robust downscaling methods account for the so-called non-Bayesian uncertainties, i.e., not complete, unobservable, or erroneous information or data. In numerous case studies in China, Ukraine, Brazil, the approaches allowed to derive local development projections of land use and land use change consistently with existing trends and expectations.

Keywords: sustainable development planning, uncertainties, robust downscaling, local-global interdependencies.

Ермольев Юрий Михайлович,
академик НАН Украины, профессор; ведущий научный сотрудник Программы прогрессивных методов системного анализа Международного института прикладного системного анализа, Лаксенбург, Австрия, e-mail: ermoliev@iiasa.ac.at.

Ермольева Татьяна Юрьевна,
PhD, научный сотрудник Программы по услугам и управлению экосистемами Международного института прикладного системного анализа, Лаксенбург, Австрия, e-mail: ermol@iiasa.ac.at.

Хавлик Петр,
PhD, научный сотрудник Программы по услугам и управлению экосистемами Международного института прикладного системного анализа, Лаксенбург, Австрия, e-mail: havlik.petr@gmail.com.

Монье Алина,
PhD, научный сотрудник Программы по услугам и управлению экосистемами Международного института прикладного системного анализа, Лаксенбург, Австрия, e-mail: mosnier@iiasa.ac.at.

Леклер Давид,
PhD, научный сотрудник Программы по услугам и управлению экосистемами Международного института прикладного системного анализа, Лаксенбург, Австрия, e-mail: leclere@iiasa.ac.at.

Фритц Стэфан,
Dr., ведущий научный сотрудник Программы по услугам и управлению экосистемами Международного института прикладного системного анализа, Лаксенбург, Австрия, e-mail: fritz@iiasa.ac.at.

Оберштайнэр Михаил,
Dr., директор Программы по услугам и управлению экосистемами Международного института прикладного системного анализа, Лаксенбург, Австрия, e-mail: oberstei@iiasa.ac.at.

Киризюк Сергей Викторович,
кандидат экон. наук, старший научный сотрудник Института экономики и прогнозирования НАН Украины, Киев, e-mail: kyryzyuk.ief@gmail.com.

Бородина Елена Николаевна,
доктор экон. наук, член-кор. НАН Украины, профессор, заведующая отделом Института экономики и прогнозирования НАН Украины, Киев, e-mail: olena.borodina@gmail.com.