

МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

DOI: 10.15838/esc.2019.6.66.4

УДК 332.144; 314.9; 519.876.5; 004.94, ББК 65.04; 60.7; 65.050

© Макаров В.Л., Бахтизин А.Р., Сушко Е.Д., Сушко Г.Б.

Агент-ориентированная суперкомпьютерная демографическая модель России: анализ апробации*



**Валерий Леонидович
МАКАРОВ**

Центральный экономико-математический институт РАН
Москва, Российская Федерация, 117418, Нахимовский пр., д. 47
E-mail: makarov@cemi.rssi.ru
ORCID: 0000-0002-2802-2100; ResearcherID: I-9022-2016



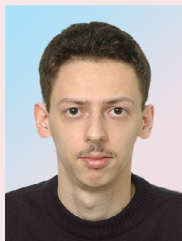
**Альберт Рауфович
БАХТИЗИН**

Центральный экономико-математический институт РАН
Москва, Российская Федерация, 117418, Нахимовский пр., д. 47
E-mail: albert.bakhtizin@gmail.com
ORCID: 0000-0002-9649-0168; ResearcherID: S-6203-2016



**Елена Давидовна
СУШКО**

Центральный экономико-математический институт РАН
Москва, Российская Федерация, 117418, Нахимовский пр., д. 47
E-mail: sushko_e@mail.ru
ORCID: 0000-0003-3565-5210; ResearcherID: E-4911-2015



**Геннадий Борисович
СУШКО**

Центральный экономико-математический институт РАН
Москва, Российская Федерация, 117418, Нахимовский проспект, д. 47
E-mail: gennady.sushko@gmail.com
ORCID: 0000-0003-3591-7276; ResearcherID: D-9438-2015

* Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научного проекта № 19-18-00240 «Суперкомпьютерные технологии в общественных науках» с использованием вычислительных ресурсов Межведомственного суперкомпьютерного центра Российской академии наук (МСЦ РАН), а также суперкомпьютера Тяньхэ-2 Национального суперкомпьютерного центра Университета оборонных технологий Китайской народной республики в г. Гуанчжоу.

Для цитирования: Агент-ориентированная суперкомпьютерная демографическая модель России: анализ апробации / В.Л. Макаров, А.Р. Бахтизин, Е.Д. Сушко, Г.Б. Сушко // Экономические и социальные перемены: факты, тенденции, прогноз. 2019. Т. 12. № 6. С. 74–90. DOI: 10.15838/esc.2019.6.66.4

For citation: Makarov V.L., Bakhtizin A.R., Sushko E.D., Sushko G.B. Agent-based supercomputer demographic model of Russia: approbation analysis. *Economic and Social Changes: Facts, Trends, Forecast*, 2019, vol. 12, no. 6, pp. 74–90. DOI: 10.15838/esc.2019.6.66.4

Аннотация. В статье представлена агент-ориентированная демографическая модель России, предназначенная для запуска на суперкомпьютерах. Используемые в модели технологии позволяют создавать искусственное общество с числом агентов до 10^9 и эффективно распараллеливать работу симулятора. Программный комплекс, созданный для реализации модели, объединяет отдельные подсистемы, написанные на языках программирования разного уровня. Что, с одной стороны, обеспечивает эффективную балансировку нагрузки между вычислительными процессами и обмен сообщениями между агентами (реализовано на языке C++), а с другой стороны, упрощает разработку блоков модели, реализующих симуляцию демографических процессов (реализовано на C#). Демографические процессы в модели имитируются на основе действий отдельных агентов с учетом их родственных связей, которые они поддерживают, обмениваясь сообщениями. Ключевыми особенностями демографической агент-ориентированной модели являются следующие: а) динамическое изменение численности и состава популяции агентов — удаление части агентов (их «смерть») и возникновение новых («рождение»); и б) разделение действий, выполняемых на шаге имитации по этапам, в конце каждого из которых может происходить пересмотр общих параметров, относящихся к регионам или группам агентов, и/или обмен сообщениями между агентами. Модель в ходе компьютерных экспериментов прошла апробацию на реальных данных и показала высокие результаты при тестировании по следующим параметрам: а) качество воссоздания на популяции агентов возрастно-половой структуры населения как по стране в целом, так и в разрезе регионов; б) устойчивость работы модели и низкая погрешность получаемых результатов прогнозирования основных демографических показателей в сравнении с вариантами официального прогноза Росстата; в) эффективность распараллеливания программного кода при запуске на суперкомпьютерах. Модель является базовой для разрабатываемой комплексной региональной имитационной модели, однако может быть полезна как самостоятельный инструмент прогнозирования.

Ключевые слова: агент-ориентированное моделирование, имитация демографических процессов, суперкомпьютерные технологии, применение графовой декомпозиции METIS, демографический прогноз для России.

Введение

Данная статья посвящена разработке агент-ориентированной демографической модели, предназначенной для создания искусственного общества с большим числом агентов-людей, и является продолжением исследования, представленного в [1]. Агент-ориентированные модели (АОМ) — это класс моделей, основанных на имитации поведения отдельных агентов, которые способны действовать самостоятельно в соответствии со своими интересами и с учетом возможностей, предоставляемых им окружающей средой (в том числе другими агентами). Демографические процессы, такие как смертность, рождаемость и миграция, являются классическими примерами процессов, формирующихся снизу вверх, так как решения принимаются на уровне отдельных индивидуумов, а общие показатели складываются в результате агрегирования действий этих индивидуумов. Поэтому демография —

это одно из направлений применения агент-ориентированного подхода, широко представленное в литературе.

Здесь достаточно упомянуть наиболее характерные работы, опубликованные за рубежом в последние годы.

Агент-ориентированные модели социальных взаимодействий и демографического поведения [2–3], в которых рассматриваются различные составляющие демографической системы, такие как процессы создания браков, изменение рождаемости и др. В работах, в том числе, исследуются различия в поведении людей, связанные с их принадлежностью к разным типам культуры и соответствующими различиями в репродуктивном поведении.

Искусственные модели популяций [4–5], в которых рассматривается АОМ с агентами сложной структуры и большим числом состояний, позволяющая прогнозировать демографиче-

скую динамику на различных уровнях — от домашних хозяйств до всего населения Великобритании.

Наиболее полно направления применения АОМ для имитации множества демографических процессов, начиная с образования семейных пар и влияния социальных норм на рождаемость и заканчивая принятием людьми решений о перемене места жительства, — представлены в книге «*Агент-ориентированная компьютерная демография: использование моделирования для улучшения понимания демографического поведения*» [6].

Очевидно, что для полноценной имитации демографических процессов на уровне большой страны необходимо создавать в модели популяции агентов большой численности, что превращает АОМ в многоагентную (мультиагентную) систему [7] и требует для проведения расчетов использования высокопроизводительной техники и специализированного программного обеспечения.

В указанной статье [1] отмечались те специфические особенности многоагентных моделей, предназначенных для имитации динамики больших социально-экономических систем, которые важны с точки зрения организации распараллеливания их работы:

- агенты популяции имеют привязку к тому или иному региону (юрисдикции), а также к координатам на карте региона и страны (имеют место жительства);
- численность и пространственное распределение популяции агентов меняются в течение периода имитации (симуляции), так как в ходе компьютерного эксперимента агенты могут не только самостоятельно мигрировать из региона в регион, но и уничтожаться и создавать новых агентов;
- агенты обладают «социальными связями» с другими агентами, с которыми они с разной интенсивностью обмениваются сообщениями, причем эти связи могут образовываться агентами из разных юрисдикций динамически в течение периода имитации;
- и место жительства агентов, и их связи с другими агентами существенным образом влияют на поведение агентов при имитации различных процессов (выбор места жительства, образование семьи, рождение детей и др.).

В статье также на примере демографической агент-ориентированной модели России обосновывался выбор средств программирования многоагентных АОМ, позволяющих реализовать авторский подход к созданию достаточно реалистичных симуляций демографических процессов, хорошо масштабируемых для популяции агентов с численностью, равной численности населения нашей страны.

Отметим, что в последние несколько лет появились специализированные программные среды для построения агентных моделей с функцией автоматического или полуавтоматического распараллеливания программного кода с последующим запуском на суперкомпьютерах.

Repast for High Performance Computing (RepastHPC) — наиболее известное программное обеспечение (ПО), разработанное для проектирования АОМ с целью их реализации в высокопроизводительных средах [8].

Высокопроизводительное ПО для построения крупномасштабных АОМ — Пандора (Pandora) — разработано в суперкомпьютерном центре Барселоны (Barcelona Supercomputing Centre) и предоставляет полную поддержку геоинформационных систем (ГИС), что важно в тех случаях, когда для функционирования моделей необходима географическая привязка агентов [9].

Среда построения агентных моделей ABM++, первая версия которой появилась в 2009 г. как результат модернизации инструмента, разработанного в 1990–2005 гг. в Лос-Аламосской национальной лаборатории в процессе построения крупномасштабных АОМ [10].

SWAGES — расширяемая распределенная среда для крупномасштабного агент-ориентированного моделирования — разработка ученых из Университета Тафтса (Медфорд, Массачусетс, США), предоставляет возможность автоматического распараллеливания программного кода и поддержку нескольких языков программирования, а также подключение плагинов для визуализации, статистического анализа и автоматической обработки ошибок [11].

CyberGIS Toolkit — набор свободно распространяемых программных компонентов с открытым исходным кодом для пространственного анализа и моделирования [12].

HPAVM – иерархическая среда параллельного моделирования, предназначенная для разработки сложных агентных моделей для исследования крупномасштабных задач, связанных с геопространственным моделированием [13].

Пакет **D-MASON** позволяет реализовывать АОМ в распределенной среде, увеличивая их производительность, при этом обеспечивая обратную совместимость с базовой средой **MASON**. Работа **D-MASON** основана на парадигме master/slave (ведущий-ведомый), при использовании которой главное приложение разделяет моделируемое пространство на части и распределяет рабочую нагрузку по ведомым процессам, каждый из которых задействует один или несколько логических процессоров (Logical Processors, LP) в соответствии с их вычислительными возможностями. Основные задачи, решаемые с помощью **D-MASON**: распределение выполняемой работы, балансировка нагрузки, связь между процессами, синхронизация и воспроизводимость [14]

В Аргоннской национальной лаборатории (национальный исследовательский центр Министерства энергетики США) разработано ПО для построения АОМ, в основном используемых для симуляции транспортных потоков – **POLARIS**. Основные утилиты разработанного пакета: (1) модуль, отвечающий за параллельную обработку событий; (2) модуль, реализующий межпроцессный обмен; (3) библиотека для визуализации; (4) библиотека для ввода-вывода данных и др. По своей сути **POLARIS** – это набор низкоуровневых библиотек, представляющих пользователю удобный программный интерфейс, а также среду выполнения, облегчающую написание программного кода. В разработанных с помощью **POLARIS** транспортных симуляторах используется большой объем данных, которые часто обрабатываются одновременно различными программными модулями. В этой связи, в параллельном режиме используются директивы для программирования многопоточных приложений с общей памятью [15].

Исследователи из Автономного университета Барселоны разработали инструмент для распараллеливания АОМ – **Care HPS** (High Performance Simulation), позволяющий в автоматическом режиме решать задачи распределе-

ния выполняемого кода, балансировки вычислительной нагрузки, связи и синхронизации. **Care HPS** поддерживает как интерфейс передачи сообщений MPI, так и технологию OpenMP и содержит в себе несколько компонент, реализованных на языке C++. Пользователи решают задачу проектирования модели (в том числе с использованием готовых функциональных элементов управления), а всю работу по распределению агентов по процессорам, синхронизации процессов и т.д. выполняет фреймворк. В настоящее время разработчики **Care HPS** используют этот фреймворк в проекте, направленном на прогноз распространения лихорадки денге [16].

Число специализированных программных продуктов для реализации агентных моделей на суперкомпьютерах постоянно растет. Помимо описанных выше, следует упомянуть также **MUSE** [17], **LUNES** [18], **MASS** [19] и др.

Однако все перечисленные средства изначально разрабатывались для решения специализированных задач, и хотя в дальнейшем они значительно расширились до уровня фреймворков широкого профиля, тем не менее сохранили определенное «наследство» в виде недостаточно эффективных алгоритмов для решения других задач.

Практический опыт их использования показал, что в конечном счете менее затратным по времени способом построения конкретной агентной модели для последующего запуска на суперкомпьютере является ее изначальная реализация с «нуля». Тем самым можно значительно повысить эффективность распараллеливания программного кода и, кроме того, создать задел для дальнейшего ее усложнения с минимальными потерями в производительности.

Программная реализация суперкомпьютерной демографической АОМ России

С учетом выводов из приведенного выше анализа существующих программных средств для распараллеливания программного кода нами было разработано ПО, обеспечивающее:

1. Масштабируемость работы АОМ на множестве узлов вычислительного кластера.
2. Эффективность многопроцессорной работы АОМ за счет использования системной библиотеки MPI, которая обычно установлена на каждом кластере, настроена на его сетевую

систему и обеспечивает максимальную пропускную способность и минимальные задержки передачи данных. Важно также и другое свойство этой библиотеки: она доступна для использования на обычных персональных компьютерах, что позволяет применять без специальной доработки одну технологию и при разработке, и при проведении расчётов как на персональном компьютере, так и с использованием суперкомпьютеров.

3. Имитацию эволюции внутреннего состояния агентов, формирование постоянных и временных связей между ними и обмен сообщениями, а также появление и исчезновение агентов в системе.

4. Динамический механизм балансировки нагрузки между вычислительными узлами в ходе работы модели, то есть перенесение агентов с более загруженных процессоров на другие с сохранением их состояния и связей (дисбаланс возникает в ходе работы модели в связи с неравномерным исчезновением агентов, привязанных к разным процессорам, а также с появлением новых агентов).

5. Упрощение разработки АОМ и дальнейшего ее развития за счет отделения блоков, реализующих систему распараллеливания с балансировкой нагрузки между вычислительными процессами и обмена сообщениями между агентами, построенных на C++, от тематических блоков, реализующих симуляцию демографических процессов, которые написаны на языке программирования высокого уровня C#.

В работе [1] было показано, каким образом в модели на основе исходных данных о численности населения по регионам, считываемых из таблицы в формате CSV, а также информации о геометрии регионов, получаемой из карты России в растровом формате, строится прямоугольная сетка, каждая ячейка которой привязывается к определённому пикселю картинки. Поскольку каждый регион закрасен своим цветом, при считывании этого изображения устанавливалась связь ячеек прямоугольной сетки с номерами регионов и другими их характеристиками. После чего осуществлялось случайное распределение числа агентов-жителей по ячейкам сетки с заданным общим числом агентов в системе и регионе, а также расчёт декомпозиции ячеек сетки по процессорам.

Для расчёта декомпозиции сетки использовался графовый алгоритм METIS [20] с заданием весов (вариант METIS_PartGraphRecursive). Алгоритм METIS принимает на вход граф, заданный через матрицу связей в формате CSR [21] и массив весов узлов графа, а возвращает оптимальное распределение графа на заданное число частей с минимизацией связей между ними. С использованием данного алгоритма было получено распределение исходной системы по процессорам. Важно то, что такое распределение необходимо производить до создания агентов, поскольку последние должны исходно быть правильно распределены между узлами кластера. При этом численность популяции агентов и количество используемых процессоров являются параметрами модели, значения которых задаются пользователем в ходе компьютерных экспериментов.

Алгоритмы, реализующие в модели эти процессы, показаны в статье [1], а в данной работе стоит подробнее остановиться на методах обеспечения правдоподобия создаваемого искусственного аналога моделируемой системы.

Первое, что необходимо в этой связи, — это установка стартового состояния, соответствующего имеющимся данным официальной статистики за базовый год, а второе — это достаточно реалистичная имитация процессов воспроизводства населения. В данной работе мы рассматриваем имитацию процессов рождаемости и смертности населения. Имитация миграции населения потребует ввода дополнительной исходной информации и добавления соответствующих процедур, идеология которых была представлена в статье [22], что, однако, не повлияет на процедуры распараллеливания работы АОМ.

Исходной информацией для модели служат следующие данные Росстата для базового года.

- На уровне страны в целом:
 - распределение населения по полу и возрасту (возрастно-половая пирамида), тыс. чел.;
 - коэффициенты смертности (на 1000 чел. населения), дифференцированные по полу и возрасту;
 - возраст выхода на пенсию для мужчин и женщин по годам переходного периода, соответствующего пенсионной реформе 2018 года.

- На уровне отдельных регионов:
 - численность населения, тыс. чел.;
 - доля населения моложе трудоспособного возраста, %;
 - доля населения в трудоспособном возрасте, %;
 - доля населения старше трудоспособного возраста, %;
 - суммарный коэффициент рождаемости;
 - распределение рождений по возрасту матерей (доли рождений, приходящихся на матерей из когорт в пятилетних возрастных интервалах в пределах репродуктивного возраста: 15–19; 20–24; 25–29; 30–34; 35–39; 40–44 и 45–49 лет), в %.

Установка стартового состояния, соответствующего имеющимся данным официальной статистики

В начале работы модели, после считывания исходных данных, масштабирования заданной численности агентов по регионам и создания рассчитанного количества агентов в ячейках, для каждого агента необходимо определить значения его индивидуальных свойств, связанных с имитируемыми процессами воспроизводства населения. Такими свойствами, в соответствии с используемыми алгоритмами имитации, являются: возраст агента, пол, максимальное желаемое число детей в семье и число уже рожденных детей. Кроме того, агент «помнит» свои семейные связи, для чего служат его индивидуальные коллекции (списки): коллекция родителей; детей; братьев и сестер; других родственников.

Распределение значений возраста и пола агентов реализовано таким образом, чтобы максимально точно воспроизвести заданную в исходных данных возрастную-половую структуру населения как по стране в целом, так и на уровне отдельных регионов. Для этого производится дальнейшее масштабирование полученных значений численности агентов каждого региона:

а) по долям основных возрастных групп населения в каждом регионе: моложе трудоспособного; трудоспособного и старше трудоспособного возраста (с учетом заданных значений для базового года возраста выхода на пенсию женщин и мужчин), а затем

б) по долям каждой возрастной когорты в своей возрастной группе.

Полученные значения долей от общего числа агентов в регионе используются в качестве вероятности выпадения того или иного возраста для агента, принадлежащего к данному региону. Для проведения подобного масштабирования и получения значения возраста для каждого агента был разработан специализированный вспомогательный модуль. Пол агента также определяется вероятностным способом с учетом соотношения полов для полученной возрастной когорты.

В модели максимальное желаемое число детей в семье – это случайная величина, принимающая значение от одного до семи с заданным бета-распределением, смещенным влево (максимум приходится на двоих детей). Для определения конкретного значения желаемого числа детей для каждого агента был также разработан специализированный вспомогательный модуль.

После распределения свойств пола и возраста наступает этап установки родственных связей между агентами. В первую очередь, для каждого агента из коллекции агентов того же региона выбирается «мать» – агент-женщина с определенным случайным образом возрастом и с числом детей меньше максимально желаемого. Выбор возраста агента-матери осуществляется на основе использования заданного в исходных данных распределения рождений по возрасту матерей, после чего происходит увеличение числа детей у выбранного агента-матери, а также взаимное занесение агентом-ребенком, агентом-матерью и агентами-родственниками матери новых родственников в соответствующие коллекции.

Имитация процессов воспроизводства населения на каждом шаге работы модели

Тестовый пример, представленный в работе [1], был разработан для оценки влияния числа используемых процессоров на эффективность распараллеливания работы модели. Еще одним параметром тестовой модели, влияние которого на эффективность работы модели оценивалось в ходе экспериментов, было число связей агентов, с которыми они обменивались сообщениями на каждом шаге, причем эти связи устанавливались случайным образом. Кроме того, состав популяции агентов был постоянен.

В полной версии демографической АОМ России, которую мы представляем в настоящей

работе, связи между агентами устанавливаются на основе родства, и их значительно меньше, чем в тестовом примере, что сокращает количество пересылаемых на шаге сообщений. С другой стороны, при имитации процессов воспроизводства населения на первый план выходят два явления: а) динамическое изменение состава популяции агентов – удаление части имеющихся агентов (их «смерть») и возникновение новых («рождение» агентов); б) разделение действий на шаге по этапам, в конце каждого из которых может потребоваться пересмотр общих параметров, относящихся к регионам или группам агентов, и/или обмен сообщениями между агентами. Наличие подобных этапов означает появление на шаге точек синхронизации, то есть приостановки работы части процессоров, которые уже закончили обработку размещенных на них агентов и ожидают окончания работы всех остальных процессоров.

Для реализации динамического добавления и удаления агентов в системе потребовался переход от одинарного (сквозного) индекса агентов к двойному. Теперь каждый агент характеризуется номером ячейки, в которой он находится, и своим номером в этой ячейке. При этом добавление агентов приводит к тому, что в ячейке увеличивается счетчик числа агентов и его значение используется для расчета индекса агента. В процессе удаления агента происходит пересчет индексов остальных агентов в рамках его ячейки, а также удаляемый агент должен быть исключен из коллекций всех агентов, с которыми он был связан родственными узами, для чего он рассылает им сообщения.

В результате при организации имитации действий на шаге пришлось отказаться от использования региональных коллекций возрастных когорт, так как поддержание их в актуальном состоянии при удалении агентов и с переходом на двойной индекс оказалось очень затратным по времени. Отказ от коллекций повлек изменение алгоритмов имитации рождаемости. Так, вместо выбора агента-женщины, которая на шаге должна родить ребенка, из коллекции женщин заданного возраста (что делалось для воспроизводства наблюдаемого распределения рождений по возрасту матери), рассчитываются соответствующие возрастные

вероятности рождения ребенка. Эти вероятности рассчитываются для каждого региона исходя из общей численности в нем женщин репродуктивного возраста, регионального суммарного коэффициента рождаемости, распределения рождений по возрасту матери, а также численности женщин каждого возраста, для которых число рожденных детей не достигло желаемого максимума.

В итоге действия на шаге работы модели были поделены на стадии, реализующие этапы имитации демографических процессов. Каждая стадия может относиться к объектам одного из следующих типов: вся страна, отдельный регион, отдельный агент. В модели реализованы следующие стадии, относящиеся к различным этапам:

1. Этап и стадия подготовки:

- Переход к следующему году. Обнуление счетчиков, связанных с имитацией рождаемости, на уровне всей страны и на уровне регионов.

2. Этап вымирания агентов:

- **Стадия рассылки «черной метки» (на уровне агентов):** маркировка агента в случае выпавшей ему в соответствии с возрастно-половым коэффициентом смертности «черной метки» (смерти). В противном случае – увеличение на единицу возраста агента.

- **Стадия пересчета индексов агентов (на уровне всей страны):** расчет новых индексов для уцелевших агентов в тех ячейках, в которых намечены удаления агентов.

- **Стадия замены индексов (на уровне агента):** посылка всем связанным агентам обновленного индекса агента; старый индекс заменяется на новый во всех коллекциях.

- **Стадия удаления агентов (на уровне всей страны):** удаление промаркированных агентов из коллекций.

3. Этап создания новых агентов:

- **Стадия расчета вероятностей рождений (на уровне всей страны):** расчет вероятностей рождений для женщин разных возрастов (по регионам);

- **Стадия рождения агентов (на уровне агента):** создание дочернего агента в случае выпадения события рождения ребенка. Установка родственных связей.

4. Этап завершения шага имитации:

• **Стадия сбора статистики (на уровне всей страны): сбор статистики по регионам и по стране в целом и сохранение результатов во внешних файлах.**

По окончании каждой стадии проверяется очередь сообщений, и если она не пуста, то происходит рассылка сообщений адресатам.

Анализ результатов экспериментов с использованием демографической АОМ России

1. *Анализ соответствия стартового состояния модели исходным данным, а также устойчивости воссоздания заданных демографических параметров*

Параметрами модели являются численность популяции агентов и количество используемых процессоров, а стартовое распределение агентов по регионам и процессорам, так же как и назначение возраста и пола отдельным агентам, осуществляется с помощью вероятностных распределений, устойчивость которых зависит от числа испытаний. Поэтому следовало оценить устойчивость распределения свойств агентов в зависимости от значений параметров модели.

В *таблице 1* приводится сравнение результатов распределения свойств агентов, полученных в ходе модельных экспериментов, с заданным распределением населения по полу и возрасту (модельные данные о распределении агентов с помощью масштабирования переводятся в численность населения в тысячах человек). В качестве исходной информации использовались данные Росстата¹. Эксперименты проводились на персональном компьютере (то

есть для одного процессора) с целью оценить минимальный размер численности популяции агентов, необходимый для приемлемого совпадения с заданной в исходных данных возрастно-половой пирамидой населения России. Кроме того, в *таблице 1* приводятся данные об устойчивости полученного распределения агентов по полу и возрасту, оцениваемого по отклонению от средних по экспериментам значений.

Из данных *таблицы* видно, что при увеличении численности агентов с одного до двух миллионов происходит существенное улучшение анализируемых статистических показателей. Так, среднее квадратичное отклонение от заданного распределения по возрастам уменьшилось в 1,38 раза для мужчин и в 1,43 раза для женщин, а показатель, характеризующий устойчивость получаемого в ходе экспериментов распределения, улучшился в 1,93 раза. При дальнейшем увеличении численности популяции агентов до трех и четырех миллионов эти показатели продолжали улучшаться, но не столь стремительно. В итоге при численности агентов четыре миллиона среднее квадратичное отклонение снизилось почти вдвое по сравнению с базовым (в 1,75 и 1,82 раза для мужчин и женщин соответственно), а отклонение от средних по экспериментам значений — более чем втрое (в 3,38 раза). Разброс значений последнего показателя в $\pm 1,4\%$, который достигается при численности агентов два миллиона, можно считать приемлемым, и именно такую численность агентов можно признать минимально допустимой для получения релевантных результатов экспериментов на уровне всей страны.

Таблица 1. Анализ адекватности и устойчивости воссоздания в модели возрастно-половой пирамиды в зависимости от численности агентов

Численность агентов, млн.	Среднее квадратичное отклонение от заданного распределения по возрастам, тыс. чел.		Отклонение от средних по экспериментам значений по возрасту и полу, %
	Мужчины	Женщины	
1	11,0	12,0	$\pm 2,7$
2	8,0	8,4	$\pm 1,4$
3	7,6	7,5	$\pm 1,1$
4	6,3	6,6	$\pm 0,8$

Источник: составлено по результатам компьютерных экспериментов с демографической АОМ России.

¹ Демографический ежегодник России. 2017: стат. сб. / Росстат. М., 2017.

Если же оценивать необходимую численность агентов на региональном уровне, то следует исходить из получаемой численности агентов для самых малонаселенных субъектов РФ. Так, население Магаданской области составляет всего 144,1 тысяч человек, и при численности агентов в модели два миллиона на долю этого региона приходится менее двух тысяч агентов. Такого количества явно недостаточно для обеспечения устойчивости распределения свойств агентов – жителей этого региона. Например, устойчивость бета-распределения, с помощью которого для агентов определяется желаемое число детей и устанавливаются связи матерей с детьми, достигается при численности агентов 10 тысяч. Исходя из этих соображений требования к общей численности агентов должны возрасти как минимум в пять раз.

Однако все последующие эксперименты ставились для оценки устойчивости показателей на уровне всей страны.

2. Анализ устойчивости работы модели

Следующая серия экспериментов представляла собой имитацию процессов воспроизводства населения России на двадцать шагов (лет) вперед при численности популяции агентов два миллиона и постоянных значениях показателей смертности и рождаемости. Для России в целом были получены прогнозы динамики следующих показателей:

- численность населения, тыс. чел.;
- распределение населения по основным возрастным группам, %;
- общий коэффициент рождаемости на 1000 человек населения;
- общий коэффициент смертности на 1000 человек населения.

Целью экспериментов было оценить устойчивость полученных результатов моделирования. Данные *таблицы 2*, в которой приводятся границы отклонений основных выходных показателей модели от средних значений по годам прогнозного периода, позволяют оценить устойчивость этих показателей (указаны максимум и минимум по всему периоду прогнозирования).

Из данных таблицы видно, что такие показатели, как численность населения и доля населения в трудоспособном возрасте, демонстрируют удивительную устойчивость с разбросом полученных в ходе экспериментов значений порядка $\pm 0,1\%$. Следующий по величине отклонения – общий коэффициент смертности (отклонение порядка $\pm 1\%$), а наибольший разброс демонстрирует общий коэффициент рождаемости ($\pm 2\%$). Таким образом, даже при численности агентов два миллиона устойчивость работы модели можно признать удовлетворительной.

3. Анализ эффективности распараллеливания работы модели

Следующим этапом тестирования демографической АОМ стала оценка эффективности распараллеливания ее работы на суперкомпьютере. В тестовом примере производился расчет десяти шагов и бралось среднее значение времени для шага. В *таблице 3* приведены результаты запуска модели на различном числе процессоров для случаев двух и восьми миллионов агентов, а на *рисунке 1* показаны соответствующие кривые ускорения работы модели в зависимости от числа процессоров, где число процессоров по оси абсцисс растет по степеням двойки.

Таблица 2. Анализ устойчивости результирующих показателей модели в течение всего прогнозного периода

Показатель	Отклонение от средних по экспериментам значений по годам прогнозного периода, %	
	нижняя граница	верхняя граница
Численность населения, тыс. чел.	-0,07	0,06
Доля населения в трудоспособном возрасте, %	-0,12	0,12
Общий коэффициент рождаемости на 1000 чел.	-2,19	1,57
Общий коэффициент смертности на 1000 чел.	-1,02	0,88

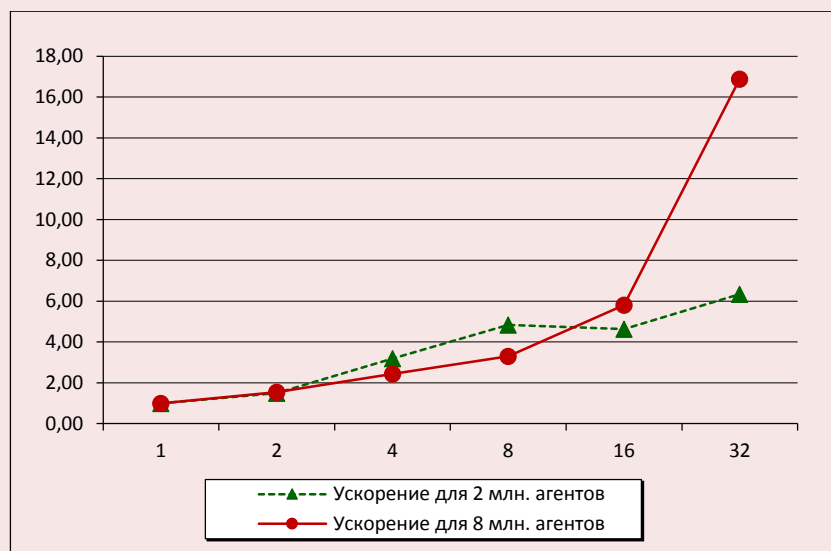
Источник: составлено по результатам компьютерных экспериментов с демографической АОМ России.

Таблица 3. Оценка ускорения при параллельном расчёте в зависимости от числа процессоров

Число процессоров	Время расчета для 2М, с	Ускорение для 2М, раз	Время расчета для 8М, с	Ускорение для 8М, раз
1	3,029	1,00	17,211	1,00
2	2,016	1,50	11,064	1,55
4	0,948	3,19	7,053	2,44
8	0,625	4,84	5,215	3,30
16	0,653	4,64	2,962	5,81
24	0,499	6,07	2,245	7,66
32	0,476	6,35	1,019	16,88
48	0,383	7,90	1,395	12,33

Источник: составлено по результатам компьютерных экспериментов с демографической АОМ России.

Рис. 1. Сравнение ускорения при параллельном расчёте в зависимости от числа процессоров при разной численности агентов, раз



Источник: рассчитано по результатам компьютерных экспериментов с демографической АОМ России.

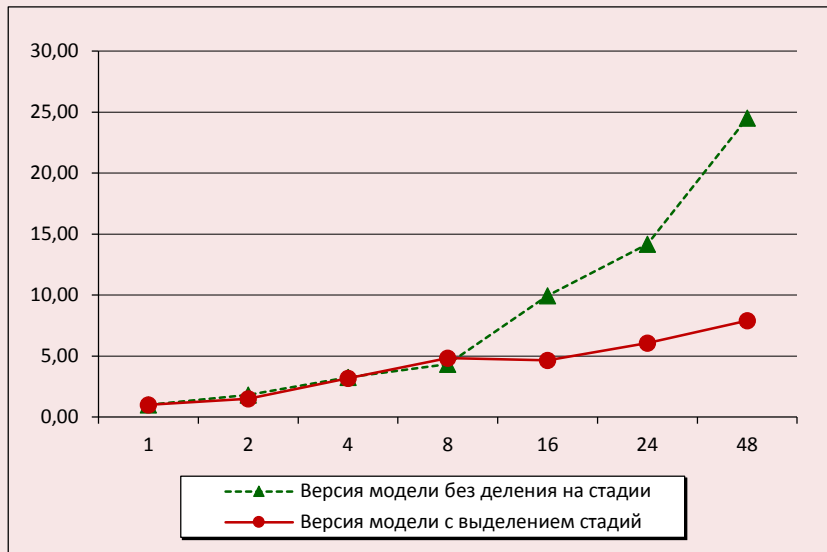
Интересно сопоставить полученную зависимость ускорения работы модели от числа процессоров с кривыми известного закона Амдала [23], который связывает максимальное достижимое ускорение с числом процессоров и с долей параллельных вычислений. Таким образом, можно оценить, насколько удачно используемые в модели алгоритмы имитации поведения агентов настроены на распараллеливание ее работы. Сравнение с теоретически достижимым уровнем ускорения показывает, что для двух миллионов агентов снижение эффективности распараллеливания в нашем случае сопоставимо с тем, как если бы доля последова-

тельных вычислений составляла примерно 10%, а для восьми миллионов – как если бы она составляла менее 5%.

4. Сравнение показателей эффективности распараллеливания работы модели при наличии точек синхронизации на шаге имитации с тестовым примером

Снижение эффективности распараллеливания во многом обусловлено наличием промежуточных точек синхронизации на каждом шаге работы модели. Поэтому большой интерес представлял вопрос о том, насколько по сравнению с первой версией модели [1] ухудшились показатели эффективности распараллеливания

Рис. 2. Сравнение ускорения работы модели при параллельном расчёте в зависимости от числа процессоров для двух версий при численности агентов 2 млн.



Источник: рассчитано по результатам компьютерных экспериментов с демографической АОМ России.

ее работы вследствие появления стадий, выполняемых на каждом шаге. Сравнение двух версий показано на *рисунке 2*.

На рисунке видно, что рост числа процессоров вплоть до восьми обеспечивает практически одинаковое ускорение работы двух версий модели, но дальнейшее увеличение числа процессоров выявляет существенную разницу в эффективности распараллеливания. Так, уже для 16-ти процессоров ускорение работы модели с дополнительными точками синхронизации вдвое меньше, чем для первой версии модели (в 2,1 раза), а для 48 процессоров — втрое (в 3,1 раза).

При оценке адекватности и эффективности авторского подхода к организации распараллеливания работы модели следует сопоставлять ее с наиболее близкими аналогами. Так, в работе [24] описывается распределенная АОМ эпидемий, в рамках которой можно создавать социум с численностью агентов до 6 млрд. В модели, так же как и в представленной авторами, используется алгоритм распределения агентов между вычислительными узлами, снижающий межгрупповое взаимодействие. А в работе [25] также рассматривается под-

ход к распараллеливанию ресурсоемких АОМ, агенты которых обмениваются информацией и имеют пространственную привязку. Ускорение, полученное авторами по сравнению с последовательной версией модели, составило в среднем 20 раз. В нашем случае в тестовом примере было получено ускорение до 25 раз при 48-ми процессорах, но добавление точек синхронизации значительно снижает этот показатель. Тем не менее, общее время расчета, как видно из таблицы 3, вполне приемлемо.

5. Результаты прогнозирования и сравнение полученного прогноза с прогнозом Росстата

Убедившись в адекватности начального состояния нашей АОМ и стабильности ее работы, оценим теперь полученные результаты моделирования с содержательной точки зрения. Так, в ходе экспериментов с моделью были получены прогнозы демографических характеристик как в разрезе регионов, так и в целом по стране. В том же Демографическом ежегоднике России 2017 года² приводятся варианты прогноза

² Демографический ежегодник России. 2017: стат. сб. / Росстат. М., 2017.

(низкий, средний и высокий) до 2051 года следующих основных характеристик населения страны в целом:

- численность населения;
- численность мужчин и женщин;
- численность населения по отдельным возрастным группам;
- родившиеся, умершие и естественный прирост населения;
- суммарный коэффициент рождаемости;
- ожидаемая продолжительность жизни при рождении.

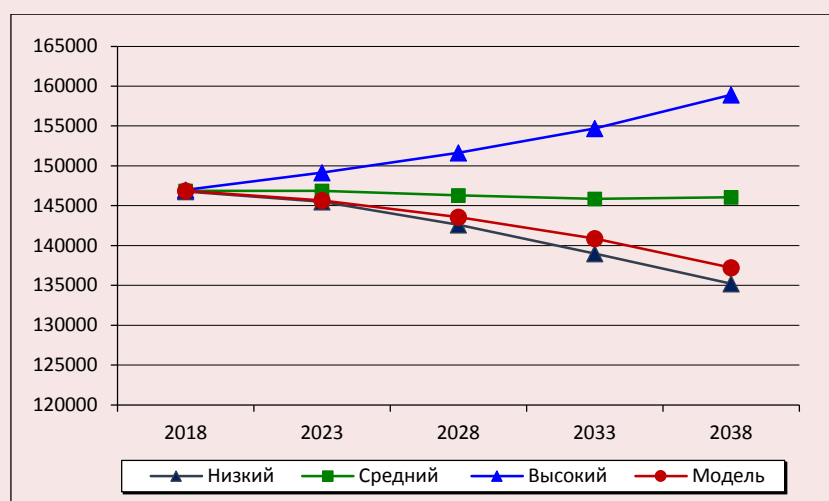
В силу того что коэффициенты смертности в экспериментах считались неизменными, так же как и суммарный коэффициент рождаемости, на основе которого и имитируется рождаемость, очевидно, что прогноз этих показателей не имеет смысла сопоставлять с прогнозами Росстата. Это же относится к такому показателю, как ожидаемая продолжительность жизни при рождении, который рассчитывается исходя из значений показателей смертности, дифференцированных по полу и возрасту.

Однако сравнить полученные с помощью модельных экспериментов прогнозы численности населения России и его структуры в разрезе отдельных возрастных групп представляется возможным и интересным. На рисунках 3 и 4

показано сравнение четырех вариантов прогноза до 2038 года: низкого, среднего и высокого варианта прогноза Росстата, а также прогноза, полученного в результате работы АОМ (взяты средние значения нескольких экспериментов при численности популяции агентов два миллиона). На *рисунке 3* видно, что модельный прогноз численности населения в течение всего периода незначительно превышает низкий вариант Росстата (на конец периода – на 1782,6 тыс. чел., или 1,3% от численности населения России), в то время как от среднего и высокого вариантов модельный отличается значительно (на -6,2 и -13,8% соответственно).

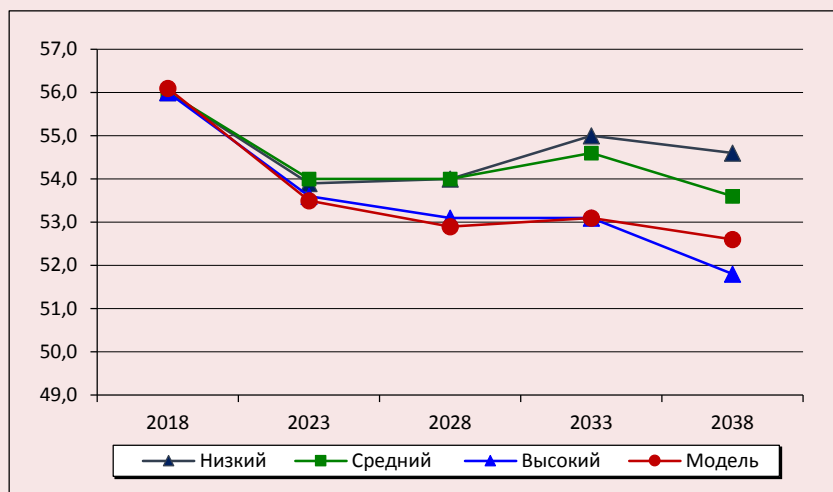
На *рисунке 4* приводится сравнение модельного прогноза доли трудоспособного населения с соответствующими вариантами прогноза Росстата. Видно, что на протяжении всего периода модельный прогноз практически совпадает с высоким вариантом Росстата, а в конце периода превышает его всего на 0,8%. Наибольшее отклонение модельного прогноза – от низкого варианта Росстата, но и оно составило всего -2%. То есть, даже при таком допущении, как неизменность коэффициентов смертности и суммарного коэффициента рождаемости, модель позволяет получать достаточно реалистичный прогноз данного показателя.

Рис. 3. Сравнение вариантов прогноза. Численность населения, тыс. чел.



Источник: рассчитано по результатам компьютерных экспериментов с демографической АОМ России и с использованием данных Демографического ежегодника России. 2017: стат. сб. / Росстат. М., 2017. Табл. 8.1.

Рис. 4. Сравнение вариантов прогноза. Доля трудоспособного населения, %

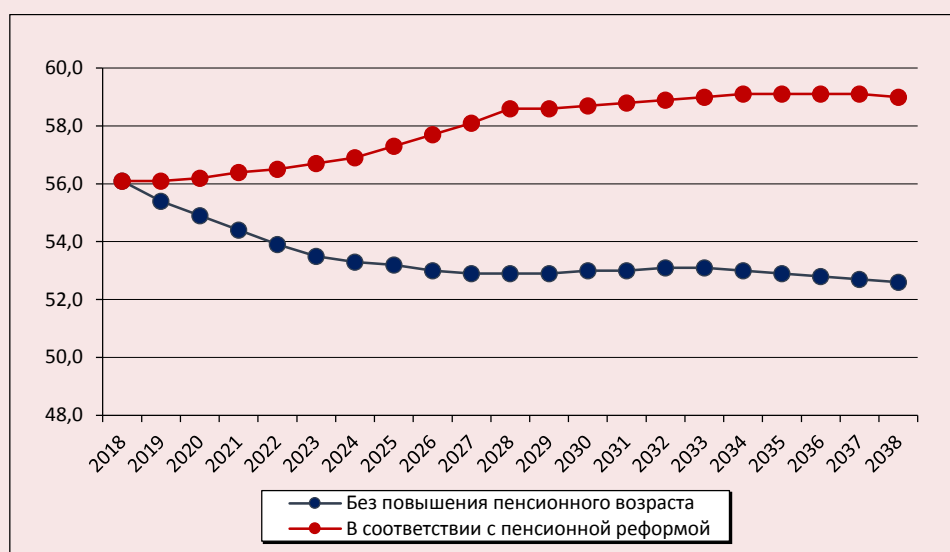


Источник: рассчитано по результатам компьютерных экспериментов с демографической АОМ России и с использованием данных Демографического ежегодника России. 2017: стат. сб. / Росстат. М., 2017. Табл. 8.3.

Следует заметить, что данный прогноз, так же как и прогнозы Росстата, был сделан в предположении о постоянстве возраста выхода на пенсию – 55 лет для женщин и 60 лет для мужчин. Однако с началом реализации пенсионной реформы в 2018 году, рассчитанной на постепенное повышение пенсионного возраста

та в течение десяти лет, ситуация изменилась, и прогноз должен это учитывать. Поэтому в следующей серии модельных экспериментов возраст выхода агентов на пенсию повышался в соответствии с условиями пенсионной реформы. Полученные результаты представлены на рисунке 5, где проводится сравнение двух

Рис. 5. Сравнение вариантов прогноза. Доля трудоспособного населения, %



Источник: рассчитано по результатам компьютерных экспериментов с демографической АОМ России.

вариантов модельных прогнозов доли трудоспособного населения: без повышения пенсионного возраста и в условиях пенсионной реформы.

На рисунке видно, как в течение десятилетнего периода реформы доля трудоспособного населения в первом варианте неуклонно снижается (с 56,1 до 52,9%), а во втором – растет (до 58,6%). В итоге в 2028 году разница уже составила 5,7 процентного пункта (что соответствует 8184,1 тыс. чел.). К концу же всего прогнозного периода доля трудоспособного населения в первом варианте снизилась до 52,6%, а во втором выросла до 59,0%, и разница между ними возросла до 6,4 процентного пункта (9189,2 тыс. чел.).

Заключение

Анализ результатов апробации представленной демографической АОМ России позволил сделать следующие выводы.

Модель в ходе всестороннего тестирования показала высокую степень устойчивости работы, несмотря на широкое применение вероятностных механизмов как при установке стартового состояния системы, так и при имитации процессов воспроизводства населения.

Используемые в модели технологии распараллеливания ее работы на множестве вычислительных узлов суперкомпьютера позволяют проводить симуляции с большим числом агентов и достигать приемлемых показателей параллельной эффективности.

Представленная демографическая модель спроектирована как базовая платформа для комплексной региональной АОМ, так как механизм разделения шагов имитации на этапы позволяет подключать к ней блоки, имитирующие любые социально-экономические процессы, в которых участвуют люди. Кроме того, реализованный механизм позволяет также указывать тип объектов, к которым этап (стадия) относится. Таким образом, модель настроена и на введение в искусственную среду объектов других типов (например, предприятий, муниципалитетов и др.) с соответствующим добавлением имитируемых процессов (социальных, экологических, экономических, политических), в которых эти объекты участвуют. Однако демографическая модель представляет и самостоятельную ценность в качестве инструмента, позволяющего в ходе экспериментов получать прогнозы демографических характеристик как в разрезе регионов, так и в целом по стране.

Литература

1. Разработка агент-ориентированной демографической модели России и ее суперкомпьютерная реализация / В.Л. Макаров, А.Р. Бахтизин, Е.Д. Сушко, Г.Б. Сушко // Вычислительные методы и программирование. 2018. Т. 19. С. 368–378. DOI: 10.26089/NumMet.v19r433.
2. Billari F.C., Prskawetz A., Diaz B.A., Fent T. The “Wedding-Ring”: an agent-based marriage model based on social interaction. *Demographic Research*, 2007, vol. 17, article 3, pp. 59–82.
3. Diaz B.A. *Agent-Based Models on Social Interaction and Demographic Behaviour (Ph.D. Thesis)*. Wien: Technische Universität, 2010. 93 p.
4. Silverman E., Bijak J., Hilton J., Cao V.D., Noble J. When demography met social simulation: a tale of two modelling approaches. *Journal of Artificial Societies and Social Simulation (JASSS)*, 2013, vol. 16 (4), article 9. Available at: <http://jasss.soc.surrey.ac.uk/16/4/9.html>.
5. Silverman E., Bijak J., Noble J., Cao V., Hilton J. Semi-artificial models of populations: connecting demography with agent-based modelling. In: Chen S.-H. et al. (Eds.). *Advances in Computational Social Science. Agent-Based Social Systems. Vol. 11*. Tokyo: Springer Japan, 2014. Pp. 177–189. DOI: 10.1007/978-4-431-54847-8_12.
6. Billari F.C., Prskawetz A. (Eds.). *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*. Heidelberg: Springer – Verlag, 2003. 210 p.
7. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. М.: Эдиториал УРСС, 2002. 352 с.

8. Collier N., North M. Parallel agent-based simulation with Repast for High Performance Computing. *Simulation*, 2012, vol. 89, no. 10, pp. 1215–1235. DOI: 10.1177/0037549712462620.
9. Wittek P., Rubio-Campillo X. Scalable agent-based modelling with cloud HPC resources for social simulations. In: *IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom). December 3-6, 2012, Taipei, Taiwan*. Pp. 355–362.
10. Roberts D.J., Simoni D.A., Eubank S. A National scale microsimulation of disease outbreaks. *Advances in Disease Surveillance*, 2007, vol. 4, no. 15.
11. Scheutz M., Connaughton R., Dingler A., Schermerhorn P. SWAGES – an extendable distributed experimentation system for large-scale agent-based alife simulations. In: *Proceedings of Artificial Life X, 2006*, pp. 412–419. Available at: <https://hrilab.tufts.edu/publications/scheutzetal06alifeswages.pdf>.
12. Shaowen W., Yan L., Anand P. Open cyberGIS software for geospatial research and education in the big data era. *SoftwareX*, 2015, no. 5. DOI: 10.1016/j.softx.2015.10.003.
13. Tang W., Wang S. HPABM: A hierarchical parallel simulation framework for spatially-explicit agent-based models. *Transactions in GIS*, 2009, no. 13 (3), pp. 315–333.
14. Cordasco G., Scarano V., Spagnuolo C. Distributed MASON: A scalable distributed multi-agent simulation environment. *Simulation Modelling Practice and Theory*, 2018, vol. 89, pp. 15–34. DOI: 10.1016/j.simpat.2018.09.002.
15. Auld J., Hope M., Ley H., Sokolov V., Xua B., Zhang K. POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. *Transportation Research Part C: Emerging Technologies*, 2016, vol. 64, pp. 101–116.
16. Borges F., Gutierrez-Milla A., Luque E., Suppi R. Care HPS: A high performance simulation tool for parallel and distributed agent-based modeling. *Future Generation Computer Systems*, 2017, vol. 68, pp. 59–73.
17. Gebre M.R. MUSE: *A parallel agent-based simulation environment (Doctoral Thesis)*. Oxford, Ohio: Miami University, 2009. 99 p.
18. D'Angelo G., Ferretti S. LUNES: Agent-based simulation of P2P systems. In: *Proceedings of 2011 IEEE International Conference on High Performance Computing & Simulation*, Istanbul, Turkey, July 2011. Pp. 593–599. DOI: 10.1109/HPCSim.2011.5999879.
19. Emau J., Chuang T., Fukuda M. A multi-process library for multi-agent and spatial simulation. In: *Proceedings of 2011 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing - PACRIM'11, Victoria, BC, Canada, August 24–26, 2011*. Pp. 369–376.
20. Karypis G., Kumar V. *METIS-unstructured graph partitioning and sparse matrix ordering system, version 2.0*. Available at: <http://dm.kaist.ac.kr/kse625/resources/metis.pdf>.
21. Tinney W., Walker J. Direct solutions of sparse network equations by optimally ordered triangular factorization. *Proceedings of the IEEE*, 1967, no. 55 (11), pp. 1801–1809.
22. Макаров В.Л., Бахтизин А.Р., Сушко Е.Д., Агеева А.Ф. Искусственное общество и реальные демографические процессы // *Экономика и математические методы*. 2017. Т. 53. № 1. С. 3–18.
23. Amdahl G.M. Validity of the single processor approach to achieving large scale computing capabilities. In: *AFIPS Conference Proceedings*, 1967, vol. 30, pp. 483–485.
24. Parker J. A flexible, large-scale, distributed agent based epidemic model. In: Henderson S.G., Biller B., Hsieh M.-H., Shortle J., Tew J.D., Barton R.R. (Eds.). *Proceedings of the 2007 Winter Simulation Conference*. Washington, D.C. December, 2007. Available at: https://www.brookings.edu/wp-content/uploads/2016/06/12_epidemicmodel_parker.pdf.
25. Gong Z., Tang W., Bennett D.A., Thill J.C. Parallel agent-based simulation of individual-level spatial interactions within a multicore computing environment. *International Journal of Geographical Information Science*, 2013, vol. 27, no. 6, pp. 1152–1170.

Сведения об авторах

Валерий Леонидович Макаров – академик РАН, доктор физико-математических наук, научный руководитель, Центральный экономико-математический институт РАН (117418, Российская Федерация, г. Москва, Нахимовский пр., д. 47; e-mail: makarov@cemi.rssi.ru)

Альберт Рауфович Бахтизин – член-корреспондент РАН, доктор экономических наук, директор, Центральный экономико-математический институт РАН (117418, Российская Федерация, г. Москва, Нахимовский пр., д. 47; e-mail: albert.bakhtizin@gmail.com)

Елена Давидовна Сушко – кандидат экономических наук, ведущий научный сотрудник, Центральный экономико-математический институт РАН (117418, Российская Федерация, г. Москва, Нахимовский пр., д. 47; e-mail: sushko_e@mail.ru)

Геннадий Борисович Сушко – кандидат физико-математических наук, старший научный сотрудник, Центральный экономико-математический институт РАН (117418, Российская Федерация, г. Москва, Нахимовский пр., д. 47; e-mail: gennady.sushko@gmail.com)

Makarov V.L., Bakhtizin A.R., Sushko E.D., Sushko G.B.

Agent-Based Supercomputer Demographic Model of Russia: Approbation Analysis

Abstract. The article presents an agent-based demographic model of Russia designed to run on supercomputers. The technologies used in the model allow researchers to create an artificial society with the number of agents up to 109 and effectively parallelize the work of the simulator. The software package designed to implement the model combines separate subsystems written in programming languages of different levels. On the one hand, this provides effective load balancing between computing processes and messaging between agents (implemented in C++), and on the other hand, this simplifies the development of model blocks that implement the simulation of demographic processes (implemented in C#). The demographic processes in the model are simulated based on the actions of individual agents, taking into account their family ties, which they maintain by exchanging messages. Key features of the demographic agent-based models are the following: a) dynamic change in the size and composition of populations of agents – removal of part of the agents (their “death”) and the emergence of new ones (“birth”); and b) separation of actions performed at the simulation step in stages, each of which can cause the revision of the general settings that are specific to regions or groups of agents, and/or exchange of messages between agents. In the course of computer experiments, the model has been tested on real data and has shown good results at testing for the following parameters: a) the quality of recreating the age-sex structure of the population for the country as a whole and in the regions with the use of the population of agents; b) the stability of the model and a low margin of error of the results of forecasting the main demographic indicators in comparison with the variants of Rosstat’s official forecast; c) efficiency of parallelization of the program code when running on supercomputers. The model is the basic one for an integrated regional simulation model that is currently being developed; however, the model can be useful as an independent forecasting tool.

Key words: agent-based modeling, simulation of demographic processes, supercomputer technologies, application of METIS graph decomposition, demographic forecast for Russia.

Information about the Authors

Valerii L. Makarov – RAS Academician, Doctor of Sciences (Physics and Mathematics), Scientific Director, Central Economic Mathematical Institute of the Russian Academy of Sciences (47, Nakhimovsky Avenue, Moscow, 117418, Russian Federation; e-mail: makarov@cemi.rssi.ru)

Al'bert R. Bakhtizin – RAS Corresponding Member, Doctor of Sciences (Economics), Director, Central Economic Mathematical Institute of the Russian Academy of Sciences (47, Nakhimovsky Avenue, Moscow, 117418, Russian Federation; e-mail: albert.bakhtizin@gmail.com)

Elena D. Sushko – Candidate of Sciences (Economics), Leading Researcher, Central Economic Mathematical Institute of the Russian Academy of Sciences (47, Nakhimovsky Avenue, Moscow, 117418, Russian Federation; e-mail: sushko_e@mail.ru)

Gennadii B. Sushko – Candidate of Sciences (Physics and Mathematics), Senior Researcher, Central Economic Mathematical Institute of the Russian Academy of Sciences (47, Nakhimovsky Avenue, Moscow, 117418, Russian Federation; e-mail: gennady.sushko@gmail.com)

Статья поступила 12.11.2019.