

УДК: 51-7

## Применение алгоритма Random Forest для построения локального оператора, уточняющего результаты расчетов в задачах внешней аэродинамики

С. В. Зимина<sup>а</sup>, М. Н. Петров<sup>б</sup>

Московский физико-технический институт (Национальный исследовательский университет),  
Россия, 141701, г. Долгопрудный, Институтский пер., д. 9

E-mail: <sup>а</sup> sofya.zimina@phystech.edu, <sup>б</sup> mikhail.petrov@phystech.edu

Получено 10.05.2021, после доработки — 21.06.2021.

Принято к публикации 22.06.2021.

При моделировании турбулентных течений неизбежно приходится сталкиваться с выбором между точностью и скоростью проведения расчетов. Так, DNS- и LES-модели позволяют проводить более точные расчеты, но являются более вычислительно затратными, чем RANS-модели. Поэтому сейчас RANS-модели являются наиболее часто используемыми при проведении практических расчетов. Но и расчеты с применением RANS-моделей могут быть значительно вычислительно затратными для задач со сложной геометрией или при проведении серийных расчетов по причине необходимости разрешения пристенного слоя. Существуют подходы, позволяющие значительно ускорить вычисления для RANS-моделей. Например, пристеночные функции или методы, основанные на декомпозиции расчетной области. Тем не менее они неизбежно теряют в точности за счет упрощения модели в пристенной области. Для того чтобы одновременно получить и вычислительно эффективную и более точную модель, может быть построена суррогатная модель на основании упрощенной модели и с использованием знаний о предыдущих расчетах, полученных более точной моделью, например из некоторых результатов серийных расчетов.

В статье строится оператор перехода, позволяющий по результатам расчетов менее точной модели получить поле течения как при применении более точной модели. В данной работе результаты расчетов, полученные с помощью менее точной модели Спаларта–Аллмараса с применением пристенной декомпозиции, уточняются на основании расчетов схожих течений, полученных с помощью базовой модели Спаларта–Аллмараса с подробным разрешением пристенной области, с помощью методов машинного обучения. Оператор перехода от уточняемой модели к базовой строится локальным образом. То есть для уточнения результатов расчета в каждой точке расчетной области используются значения переменных пространства признаков (сами переменные поля и их производные) в этой точке. Для построения оператора используется алгоритм Random Forest. Эффективность и точность построенной суррогатной модели демонстрируется на примере двумерной задачи сверхзвукового турбулентного обтекания угла сжатия при различных числах Рейнольдса. Полученный оператор применяется к решению задач интерполяции и экстраполяции по числу Рейнольдса, также рассматривается топологический случай — интерполяция и экстраполяция по величине угла сжатия  $\alpha$ .

Ключевые слова: пристенная декомпозиция, пристенные функции, вычислительная аэродинамика, случайный лес, машинное обучение, турбулентность

UDC: 51-7

# Application of Random Forest to construct a local operator for flow fields refinement in external aerodynamics problems

S. V. Zimina<sup>a</sup>, M. N. Petrov<sup>b</sup>

Moscow Institute of Physics and Technology (National Research University),  
9 Institutskiy per., Dolgoprudny, 141701, Russia

E-mail: <sup>a</sup> sofya.zimina@phystech.edu, <sup>b</sup> mikhail.petrov@phystech.edu

Received 10.05.2021, after completion — 21.06.2021.

Accepted for publication 22.06.2021.

Numerical modeling of turbulent flows requires finding the balance between accuracy and computational efficiency. For example, DNS and LES models allow to obtain more accurate results, comparing to RANS models, but are more computationally expensive. Because of this, modern applied simulations are mostly performed with RANS models. But even RANS models can be computationally expensive for complex geometries or series simulations due to the necessity of resolving the boundary layer. Some methods, such as wall functions and near-wall domain decomposition, allow to significantly improve the speed of RANS simulations. However, they inevitably lose precision due to using a simplified model in the near-wall domain. To obtain a model that is both accurate and computationally efficient, it is possible to construct a surrogate model based on previously made simulations using the precise model.

In this paper, an operator is constructed that allows reconstruction of the flow field obtained by an accurate model based on the flow field obtained by the simplified model. Spalart–Allmaras model with approximate near-wall domain decomposition and Spalart–Allmaras model resolving the near-wall region are taken as the simplified and the base models respectively. The operator is constructed using a local approach, i. e. to reconstruct a point in the flow field, only features (flow variables and their derivatives) at this point in the field are used. The operator is constructed using the Random Forest algorithm. The efficiency and accuracy of the obtained surrogate model are demonstrated on the supersonic flow over a compression corner with different values for angle  $\alpha$  and Reynolds number. The investigation has been conducted into interpolation and extrapolation both by  $Re$  and  $\alpha$ .

Keywords: approximate near-wall domain decomposition, wall functions, computational aerodynamics, random forest, machine learning, turbulence

Citation: *Computer Research and Modeling*, 2021, vol. 13, no. 4, pp. 761–778 (Russian).

## Введение

Численное моделирование турбулентных течений около стенки является естественным требованием для многих промышленных задач. При этом для получения высокой точности расчета необходимо хорошо разрешать пограничный слой. Так, при решении задачи об обтекании тела на разрешение пограничного слоя может потребоваться до 90 % расчетного времени. При проведении серийных расчетов такие затраты могут оказаться чрезмерными. Чтобы исключить вычислительную сложность, связанную с разрешением пограничного слоя, были построены специальные модели. В качестве примера таких моделей можно привести RANS-модели с использованием пристеночных функций [Grotjans, Menter, 1998; Craft et al., 2004; Kalitzin et al., 2005] или модели, основанные на методах декомпозиции расчетной области [Utyuzhnikov et al., 2018; Chikitkin et al., 2020]. Но такие модели, несмотря на их большую вычислительную эффективность по сравнению с базовыми моделями, могут существенно проигрывать им в точности. При этом некоторые специальные модели (например, метод приближенной пристенной декомпозиции (ANDD) [Petrov et al., 2020]) способны сохранить основную физику течения. Потери в точности в этом случае происходят локально. Поэтому предполагается, что с помощью методов машинного обучения возможно построить суррогатную модель, позволяющую уточнять результаты расчетов, полученные специальной моделью, на основании данных расчетов, полученных ранее для других задач в рамках более точной модели. При этом предполагается, что база данных расчетов для построения суррогатной модели была получена по расчетам из одного класса задач. Построенная таким образом суррогатная модель позволит сохранить как высокую точность вычислений, так и высокую скорость сходимости решения.

Приведем небольшую иллюстрацию в качестве демонстрации вычислительной эффективности метода при проведении вычислений на практике. В работе сделан акцент на проведении серийных расчетов. Предположим, необходимо провести серию из 10 расчетов для задачи внешнего турбулентного обтекания тела (самолета, спускаемого аппарата и т. д.) с разными параметрами набегающего потока, угла атаки или немного различной геометрией (как, например, в задачах проектирования профиля крыла). Известно, что применение специального метода ANDD сокращает время одного расчета почти втрое [Petrov et al., 2020] по сравнению с базовой моделью с хорошим разрешением пристенной области (например, модели Спаларта–Аллмараса или модели Ментера). В частности, в задаче внешнего обтекания, рассмотренной в статье, проведение одного расчета в рамках базовой модели занимало 3 часа, в рамках специальной модели — 1 час. И в первом случае, при применении только базовой модели, на проведение серии из 10 расчетов необходимо потратить 30 часов расчетного времени, с использованием специальной модели — 10 часов. То есть выигрыш во времени здесь в 3 раза. Несмотря на существенный выигрыш во времени, за счет упрощения модели приходится неминуемо терять в точности. В работе предлагается построить суррогатную модель и провести серию расчетов следующим образом: 3 расчета провести базовой моделью и еще 3 — специальной. А затем по этим расчетам провести обучение суррогатной модели, которая будет уточнять специальные расчеты в дальнейшем. На проведение этих расчетов уйдет 12 часов расчетного времени. Обучение алгоритма, используемого в статье, происходит достаточно быстро. В рамках рассматриваемой в статье задачи на это уходило примерно 20 минут. Далее оставшиеся 7 расчетов проведем только специальной моделью и уточним суррогатной. Поскольку само уточнение фактически не требует значимого времени, времязатраты здесь составят семь часов. Полное время расчетов для всей серии в этом случае составит чуть менее 20 часов. То есть экономия времени здесь в полтора раза относительно базовой модели. Но при этом точность всей серии будет сопоставима с точностью расчетов, проведенных именно базовой моделью на подробной сетке. Понятно, что выигрыш здесь может быть еще больше, если сама серия больше или используется более быстрая специальная модель. В работе используется метод ANDD в качестве специальной модели, так как дает достаточно хорошее

решение для лучшего обучения суррогатной модели. Также важно отметить, что общая методология, рассматриваемая в статье, может быть применена к любой паре выбранных специальной и базовой моделей. Построение суррогатной модели может проводиться в рамках отдельного скрипта и не зависит от выбранного расчетного солвера для задач внешней аэродинамики, а на вход программа принимает только результаты расчетов по базовой и специальной модели. На выход дает уточненный расчет или какие-то конкретные интересующие исследователя характеристики.

Применение алгоритмов машинного обучения для решения задач о разрешении турбулентных течений в газовой динамике является достаточно новым, тем не менее стремительно набирающим популярность подходом. В работе [Tracey et al., 2015] строится модель (калибрационный подход), в которой модифицируется модель Спаларта–Аллмараса, используемая в качестве замыкания уравнений RANS. С помощью алгоритма машинного обучения модифицируется источниковый член модели так, чтобы максимально точно учесть физику явления. Для этого авторами выделяется ряд признаков (например, завихренность, коэффициент вязкости) и проводится обучение на данных с высокой точностью результата (полученные из DNS-расчетов или же экспериментальных данных). Фактически в подходе производится калибровка коэффициентов в источниковом члене. Указанная работа, конечно, не единственная, в которой используется калибрационный подход, но достаточно показательная для данного направления. В работах [Parish, Duraisamy, 2016; Singh, Duraisamy, 2016; Singh et al., 2017; Ferrero et al., 2020] строится метод инверсии поля (статистический подход), который позволяет с помощью обратного моделирования построить пространственно распределенный функциональный член, уточняющий рассматриваемую турбулентную модель замыкания. Инверсия осуществляется с использованием алгоритма семплирования, который может быть вычислительно затратным. В работах [Ling, Templeton, 2015; Ling et al., 2016] представлен метод с использованием глубоких нейронных сетей для изучения тензора напряжений Рейнольдса на основе данных с высокой точностью результата. Предложенная архитектура нейронной сети использует мультипликативный слой с инвариантным тензорным базисом для встраивания галилеевой инвариантности в предсказанный тензор. Показано, что эта архитектура нейронной сети обеспечивает улучшенную точность расчетов. В работе [Hanna et al., 2020] строится модель (подход агрегирования данных), которая позволяет уточнять результаты расчетов на грубой сетке по результатам расчетов на подробной сетке, полученных для задач из того же класса. Данный подход позволяет локально предсказать ошибку, вызванную огрублением сетки, и скорректировать на нее переменные поля течения. Алгоритм машинного обучения строится на основании фактической разницы между переменными признакового пространства точного и приближенного решения. В дальнейшем для подхода агрегирования данных кажется необходимым более глубоко исследовать признаковое пространство, напрямую учитывать дифференциальные уравнения, описывающие задачу при построении модели методами машинного обучения [Raissi et al., 2015], учитывать геометрию рассматриваемой задачи [Guo et al., 2015].

Рассматриваемая в статье суррогатная модель в постановке, примененной к методу пристенной декомпозиции, построена впервые. Схожая идея была относительно успешно применена в статье [Hanna et al., 2020] для течения в каверне (lid-driven cavity) в рамках подхода CG-CFD, где в качестве более точной модели рассматривается базовая модель на подробной сетке, а в качестве уточняемой — базовая модель на грубой сетке. Подход, рассматриваемый в настоящей статье, обладает рядом преимуществ по сравнению с CG-CFD, ключевое из которых — значительно более корректное в плане разрешаемой физики поле течения, получаемое в уточняемом расчете, что должно позволить построить более точный оператор перехода к базовой модели. Так, в результате огрубления сетки для CG-CFD часть информации о явлениях, происходящих в рассматриваемой задаче, может быть потеряна. В частности, сетка, используемая для приближенного расчета, должна разрешать пограничный слой. То есть здесь возникает

естественное ограничение на степень огрубления сетки, которое может быть достаточно сильным для некоторых задач и может снизить прогнозируемое уменьшение вычислительных затрат. Также восстановление уточненного поля течения требует переинтерполяции на подробную сетку. Но в результате потерь, вызванных грубой сеткой, не только разрешаемое явление может потерять в точности, но и положение его на поле течения может оказаться значительно смещенным. Суррогатная модель, построенная в статье, лишена этих недостатков. Тесты демонстрируют способность метода приближенной пристенной декомпозиции с высокой точностью определять положения отрыва и область формирования ударной волны. Стоит отметить, что стандартные пристенные функции не могут в полной мере разрешать указанные явления. Также пристенная декомпозиция хорошо справляется с разрешением турбулентных течений. По построению пристенная декомпозиция хорошо разрешает пристенную область. Все это дополняется высокой скоростью сходимости расчетов, проводимых данным методом. К тому же для восстановления поля течения не требуется переинтерполяция, так как для проведения расчета базовой и специальной моделью в этом случае фактически используется одна и та же расчетная сетка.

Для построения суррогатной модели в исследовании используется алгоритм Random Forest (RF) [Breiman, 2001], примененный для локальной задачи: модель будет обучаться и применяться поточечно для каждой точки поля течения (фактически для каждой ячейки расчетной сетки). Для уточнения решения в каждой точке пространства в качестве признаков используются значения газодинамических переменных и их производных. RF формирует оператор перехода от числовых результатов приближенной специальной модели (ANDD для модели Спаларта–Аллмараса) к числовым результатам базовой модели Спаларта–Аллмараса.

Валидация суррогатной модели проводится путем сравнения поля течения с полем течения, получаемым базовой моделью на примере двумерной задачи сверхзвукового турбулентного обтекания угла сжатия при различных числах Рейнольдса. В частности, сравниваются такие характеристики, как трение на поверхности тела, интегральный коэффициент трения, профили скорости.

Статья организована следующим образом. В подразделе *Идея построения суррогатной модели* описан общий принцип построения суррогатной модели. Подразделы *Аэродинамический решатель*, *Уточняемая модель* и *Random Forest* описывают соответственно используемый аэродинамический солвер, уточняемую модель (ANDD) и алгоритм машинного обучения (Random Forest), используемый для построения суррогатной модели. В разделе *Результаты* рассмотрено построение уточняющего оператора на примере задачи сверхзвукового турбулентного обтекания угла сжатия и проведена валидация полученной суррогатной модели для интерполяционного, экстраполяционного и топологического случаев. Результаты суррогатной модели для каждого случая сравниваются с результатами расчетов по специальной (уточняемой) и базовой (точной) моделям. Заключение работы представлено в разделе *Заключение*.

## Методология

### *Идея построения суррогатной модели*

Схема практического применения предлагаемого в статье подхода при проведении серии из  $N$  расчетов выглядит следующим образом:

- 1)  $m$  расчетов серии проводятся базовой и специальной моделями;
- 2) на основании этих расчетов обучается суррогатная модель с помощью алгоритма машинного обучения Random Forest;
- 3) оставшиеся  $N - m$  расчетов серии проводятся быстрой специальной моделью и уточняются суррогатной моделью.



Формально проводимые по такой схеме серийные расчеты можно поделить на интерполяционные, экстраполяционные и топологические. В случае первых двух вариантов расчеты серии отличаются только параметрами набегающего потока по числу Рейнольдса и Маха. При этом предполагается, что в интерполяционном случае расчет на тесте находится в области интерполяции по параметрам набегающего потока относительно расчетов на обучении, в экстраполяционном — в области экстраполяции. В топологическом случае при тех же параметрах набегающего потока меняется геометрия обтекаемого тела. Например, для задачи обтекания угла сжатия — величина самого угла. Как можно будет видеть далее, число расчетов на первом шаге  $m$  достаточно мало для получения хорошей точности результатов после применения суррогатной модели как в интерполяционном, экстраполяционном, так и в топологическом случае. В исследовании в качестве базовой модели используется модель Спаларта–Аллмараса [Spalart, Allmaras, 1992], решение для которой получается на подробной сетке с хорошим разрешением пристенной области.

Алгоритм машинного обучения Random Forest [Breiman, 2001], используемый для построения суррогатной модели, для обучения использует фактическую разницу между переменными пространства признаков для точного решения, полученного в рамках базовой модели, и приближенного решения, полученного в рамках модели Спаларта–Аллмараса с применением метода приближенной пристенной декомпозиции (ANDD). В работе задача решается в локальной постановке, то есть для уточнения решения в каждой точке пространства используются значения признаков в этой точке. Применение алгоритма Random Forest в рамках рассматриваемой методологии можно понимать как проведение интерполяции/экстраполяции для серии расчетов, обобщенной в смысле данного алгоритма.

При выборе признаков предполагалось, что ошибка в каждой ячейке расчетной сетки зависит от значения переменной в данной и соседних ячейках, поэтому в качестве признаков используются значения переменных и их производные по координатам. В работе в качестве исследуемых переменных выбраны компоненты скорости  $v_x$  и  $v_y$ . Поэтому в качестве признаков используются значения компонент скорости в расчете, получаемые методом ANDD, и их частные производные  $\left(\frac{\partial v_x}{\partial x}, \frac{\partial v_x}{\partial y}, \frac{\partial v_y}{\partial x}, \frac{\partial v_y}{\partial y}\right)$ . Еще одним важным параметром, который повлиял на качество обучаемой модели, является расстояние от стенки  $l$ . Таким образом, вектор признакового описания имеет вид

$$X = \left( v_x, v_y, \frac{\partial v_x}{\partial x}, \frac{\partial v_x}{\partial y}, \frac{\partial v_y}{\partial x}, \frac{\partial v_y}{\partial y}, l \right). \quad (1)$$

В качестве целевой переменной, которую обучается предсказывать суррогатная модель, используется величина ошибки между базовой и приближенной моделями.

Выражение для целевой переменной получается из следующих соображений. Пристенная область является областью наибольшего интереса, так как именно по значениям компонент скорости вблизи стенки рассчитываются трение на стенке  $f_t$  и коэффициент трения  $C_f$ , а также именно вблизи стенки наблюдается самое большое расхождение результатов расчетов по базовой и приближенной моделям. Поэтому для того, чтобы увеличить значимость ошибки в пристенной области (где абсолютная величина скорости мала), применяется логарифмирование, и в качестве целевой переменной используется разность между логарифмами компоненты скорости ( $v_x$  или  $v_y$ ) для одноблочного расчета и расчета методом ANDD:

$$y = \ln(v_{\text{single block}} + a) - \ln(v_{\text{ANDD}} + a), \quad (2)$$

где  $v$  — это  $v_x$  или  $v_y$ . Константа  $a$  выбирается таким образом, чтобы для всех точек поля течения величина под логарифмом была положительной. Значение  $a$  было выбрано чуть выше максимального по модулю отрицательного значения компоненты скорости ( $v_x$  или  $v_y$ ) для всей выборки:

$$a = |\min(v)| + k,$$

где  $k$  — малое положительное число. В качестве функции потерь при обучении алгоритма Random Forest в работе использовалась среднеквадратичная ошибка (MSE), в качестве критерия останова обучения — ограничение на максимальную глубину дерева (максимальная глубина равнялась 15).

### **Аэродинамический решатель**

Все расчеты проведены на основе имеющегося кода лаборатории *Flowmodellium* [Petrov et al., 2018] для решения пространственных уравнений Навье–Стокса на неструктурированных сетках. Уравнения RANS решаются с использованием метода конечных объемов. В качестве модели замыкания используется модель Спаларта–Алламараса. Невязкие потоки получаются из решения задачи Римана о распаде разрыва, вязкие потоки и источниковые члены вычисляются с помощью аппроксимации пространственных производных. В результате система RANS сводится к системе нелинейных ОДУ. Применение неявной схемы Эйлера приводит к системе алгебраических уравнений. Данная система решается с использованием схемы LU-SGS [Петров и др., 2017; Chikitkin et al., 2018]. Для решения задач с большим количеством расчетных ячеек в программном комплексе реализована двухуровневая модель параллельных вычислений.

### **Уточняемая модель**

Для специальной (уточняемой) модели расчет проводится по методу ANDD (Approximate Near-wall Domain Decomposition). Суть метода ANDD заключается в разделении расчетной области ( $0 \leq y \leq y_e$ ) на внутреннюю ( $0 \leq y \leq y^*$ ) и внешнюю ( $y^* \leq y \leq y_e$ ) области, где  $y$  — координата в направлении нормали к стенке. При этом во внутренней (пристенной) области уравнения Навье–Стокса решаются в приближении тонкого слоя:

$$\frac{\partial}{\partial y} \left( \mu \frac{\partial U}{\partial y} \right) = f, \quad U(0) = U_0, \quad U(y_e) = U_1, \quad (3)$$

где  $f$  — некая правая часть. Интегрируя уравнение (3) дважды, можно получить граничное условие третьего рода на границе ( $y = y^*$ ) между внешней и внутренней областями:

$$U^* = \frac{\partial U}{\partial y} \Big|_{y^*} f_1 + f_2 + U_0. \quad (4)$$

Здесь

$$f_1 = \int_0^{y^*} \frac{\mu^*}{\mu} dy, \quad f_2 = - \int_0^{y^*} \frac{1}{\mu} \int_0^{y'} f dy dy', \quad \mu^* = \mu(y^*). \quad (5)$$

Граничные условия имеют одинаковый вид (4) для всех примитивных переменных, кроме нормальной скорости  $v_n$ .

Ниже представлен алгоритм расчета по методу ANDD:

- 1) инициализация поля течения во внутренней и внешней областях;
- 2) получение граничных условий третьего рода на границе между внешней и внутренней областями из уравнения пограничного слоя (3) для внутренней области;
- 3) решение полных уравнений во внешней области с граничным условием третьего рода (4) на границе между областями; данный шаг соответствует одной глобальной итерации метода LU-SGS во внешней области;
- 4) пересчет поля течения во внутренней области с использованием граничного условия Дирихле на границе между областями, полученного из решения задачи во внешней области.

Шаги 2–4 повторяются до выполнения условия сходимости.

Так как в работе в качестве исследуемых переменных выбраны компоненты скорости, опишем, как получаются граничные условия для тангенциальной скорости  $v_\tau$  и нормальной скорости  $v_n$ , более подробно.

Закон сохранения импульса в приближении тонкого слоя имеет вид

$$\frac{\partial}{\partial y} \left( \mu \frac{\partial v_\tau}{\partial y} \right) = \frac{dp}{dx}. \quad (6)$$

Интегрируя уравнение (6) дважды, можно получить граничное условие вида (4) для тангенциальной скорости на границе между областями:

$$v_\tau^* = f_1^v v_{\tau,y}^* + f_2^v, \quad (7)$$

которое может быть приближено как

$$v_\tau^* = \frac{v_{\tau,I} f_1^v + d f_2^v}{d + f_1^v}.$$

Здесь  $v_\tau^*$  — тангенциальная компонента скорости на границе между внешней и внутренней областями,  $v_{\tau,I}$  — тангенциальная компонента скорости в первой ячейке возле границы во внешней области,  $d$  — расстояние от центра первой ячейки до границы,  $f_1^v$  и  $f_2^v$  — интегралы вида (5) для тангенциальной скорости.

Когда решение во внешней области получено, профиль тангенциальной компоненты скорости во внутренней области может быть получен по следующей формуле:

$$v_\tau(y) = \frac{dv}{dy} \Big|_{y^*} \int_0^y \frac{\mu^*}{\mu(\xi)} d\xi - \int_0^y \frac{dp/dx (y^* - \xi)}{\mu(\xi)} d\xi. \quad (8)$$

Нормальная компонента скорости  $v_n$  вблизи стенки (при разложении по формуле Тейлора) имеет вид

$$v_n \approx \alpha \frac{y^2}{2} + \beta \frac{y^3}{6}.$$

Отсюда производная  $v_n'$ :

$$v_n' = \alpha y + \beta \frac{y^2}{2}.$$

Из уравнений движения получаем  $\alpha = \frac{p_x}{\mu}$ . Исключая  $\beta$ , получаем граничное условие для  $v_n$ :

$$v_n = \frac{y^*}{3} v_n' + \frac{y^{*2}}{6} \frac{p_y}{\mu}.$$

Или в первом приближении:

$$v_n = \frac{y^*}{3} v_n'. \quad (9)$$

Аппроксимируя (9) вблизи границы между областями, получаем граничное условие:

$$v_n^* = \frac{y^*}{3d + y^*} v_{n,I},$$

где  $v_{n,I}$  — нормальная компонента скорости в первой ячейке возле границы во внешней области.

Когда решение во внешней области получено, профиль нормальной компоненты скорости может быть получен по следующей формуле:

$$v_n(y) = v_n(y^*) \left( \frac{y}{y^*} \right)^2.$$



**Random Forest**

Для построения уточняющего оператора в работе используется алгоритм Random Forest [Breiman, 2001]. Random Forest — это алгоритм машинного обучения, заключающийся в использовании ансамбля решающих деревьев. Дерево решений представляет собой объединение логических условий (вида  $x_1 < a_1, x_2 < a_2, \dots, x_M < a_M$ , где  $(x_1, x_2, \dots, x_M)$  — признаки) в структуру дерева. При построении дерева в каждом промежуточном узле один из признаков сравнивается с некоторым порогом и попавшая в узел подвыборка делится на две части, каждая из которых попадает в свое поддерево. Оптимальное разбиение в каждой вершине выбирается из условия максимизации функционала качества:

$$Q(R_m) = H(R_m) - \frac{|R_l|}{|R_m|}H(R_l) - \frac{|R_r|}{|R_m|}H(R_r). \tag{10}$$

Здесь  $R_m$  — множество объектов, попавшее в данную вершину;  $R_l$  — множество объектов, попавшее в левое поддерево,  $R_r$  — множество объектов, попавшее в правое поддерево.  $H(R)$  — критерий информативности, который оценивает качество распределения целевой переменной среди объектов множества  $R$ . Чем меньше разнообразие целевой переменной, тем меньше значение критерия информативности. В качестве критерия информативности в задачах регрессии обычно используется среднеквадратичная ошибка (MSE) или средняя абсолютная ошибка (MAE). Процедура разбиения продолжается рекурсивно, при выполнении условия останова в вершине вершина объявляется листом и далее не разбивается. В качестве критерия останова может использоваться, например, ограничение на максимальную глубину дерева, максимальное число листьев, минимальное число объектов в листе. Когда дерево построено, каждому листу ставится в соответствие значение целевой функции. Для задач регрессии это может быть, например, среднее значение или медиана целевой функции для объектов в листе. Пример решающего дерева показан на рис. 1 (слева).

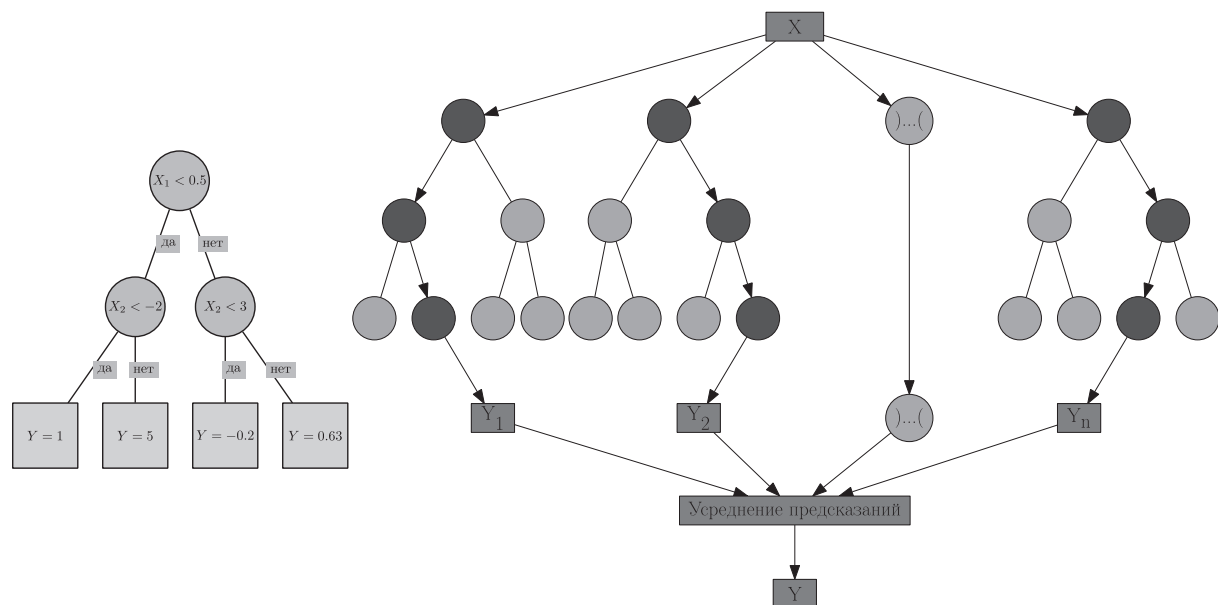


Рис. 1. Пример решающего дерева регрессии (слева) и случайного леса (справа)

К достоинствам решающих деревьев относятся хорошая интерпретируемость результатов и быстрота обучения. Значительным недостатком алгоритма является склонность к переобучению. Переобучение (overfitting) — явление, когда модель выдает недостаточно качественные

результаты на данных, которых не было в ее обучающей выборке, при хороших результатах на данных из обучающей выборки; иными словами, модель обладает плохой обобщающей способностью. В связи с этим на практике одиночные решающие деревья применяются редко; более часто используют ансамбли деревьев.

Одним из вариантов построения ансамбля решающих деревьев является случайный лес (Random Forest). Предположим, что имеются обучающая выборка, состоящая из  $n$  объектов, и множество из  $M$  признаков. Тогда алгоритм построения случайного леса из  $N$  деревьев выглядит следующим образом.

Для каждого  $k = 1, \dots, N$

- генерируется случайная выборка с повторениями размера  $n$  из исходной выборки;
- на основе сгенерированной выборки строится решающее дерево, при этом
  - в каждом промежуточном узле по заданному критерию (10) выбирается оптимальный признак, по которому проводится разбиение на два поддерева;
  - разбиение продолжается рекурсивно до выполнения критерия останова (например, максимальная глубина дерева или минимальное число объектов в листе);
  - при выборе лучшего признака при разделении сначала из исходных  $M$  признаков выбираются случайным образом  $m$  признаков, и оптимальный выбирается только среди них (обычно в задачах регрессии берется  $m = \frac{M}{3}$ ).

Итоговое предсказание значения целевой функции для объекта тестовой выборки получается усреднением предсказаний всех деревьев. Рис. 1 (справа) иллюстрирует работу алгоритма.

Оптимальное число деревьев в случайном лесе в данной работе определялось по минимизации ООВ (Out of bag) ошибки. ООВ-ошибка — это ошибка каждого решающего дерева на тех объектах, которые не попали в его обучающую подвыборку, усредненная по всем деревьям. Число деревьев в случайном лесе для каждого сценария выбиралось по минимальной ООВ-ошибке. На рис. 2 показана зависимость ООВ-ошибки от числа деревьев (на примере задачи интерполяции по числу Рейнольдса  $Re$ ). В качестве функции потерь при обучении случайного леса использовалась среднеквадратичная ошибка (MSE).

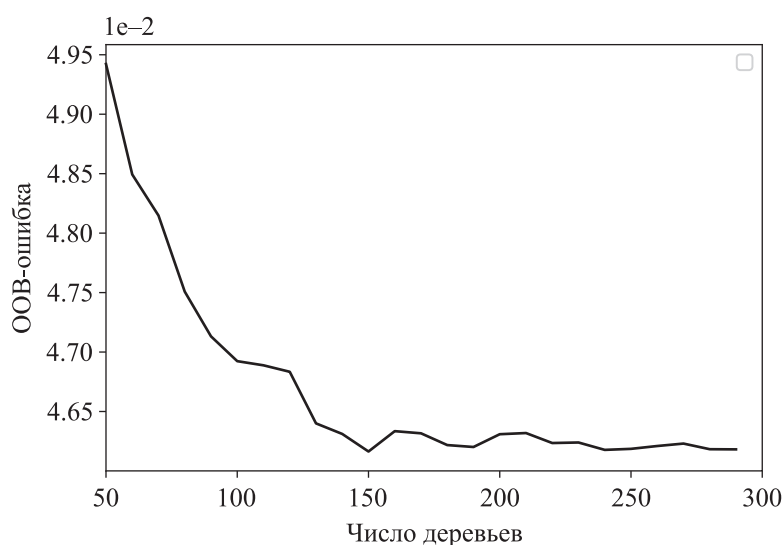


Рис. 2. Зависимость ООВ-ошибки от числа деревьев

## Результаты

Оператор перехода от уточняемой модели к базовой строится на примере двумерной задачи сверхзвукового турбулентного обтекания угла сжатия с числом Маха  $M = 1.5$  при различных числах Рейнольдса  $Re$  в диапазоне от  $9 \cdot 10^6$  до  $12 \cdot 10^6$  и величине угла сжатия  $\alpha$  в диапазоне от  $128^\circ$  до  $140^\circ$ . В качестве исследуемых переменных выбраны компоненты скорости  $v_x$  и  $v_y$ , по значениям которых также считается трение на стенке  $f_\tau$  и коэффициент трения  $C_f$ . На рис. 3 показана расчетная сетка для случая  $\alpha = 132^\circ$ . Расчетная сетка сгущается к стенке, размер первой пристенной ячейки равен  $10^{-6}$  м. Для модели ANDD граница между внутренней и внешней областью находится на расстоянии от стенки, соответствующем  $y^+ \approx 100$ . На рис. 4 показаны поля компоненты скорости  $v_x$  для задачи с  $Re = 10^7$ ,  $\alpha = 132^\circ$ , рассчитанные с помощью базовой одноблочной модели и менее точной модели ANDD.

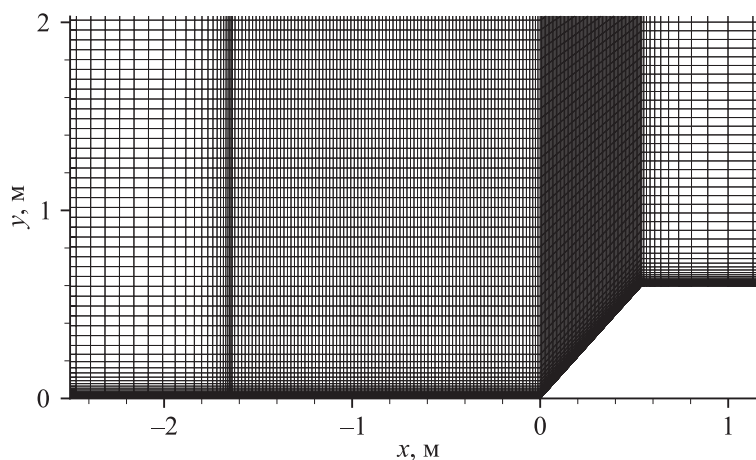


Рис. 3. Расчетная сетка для задачи с  $\alpha = 132^\circ$

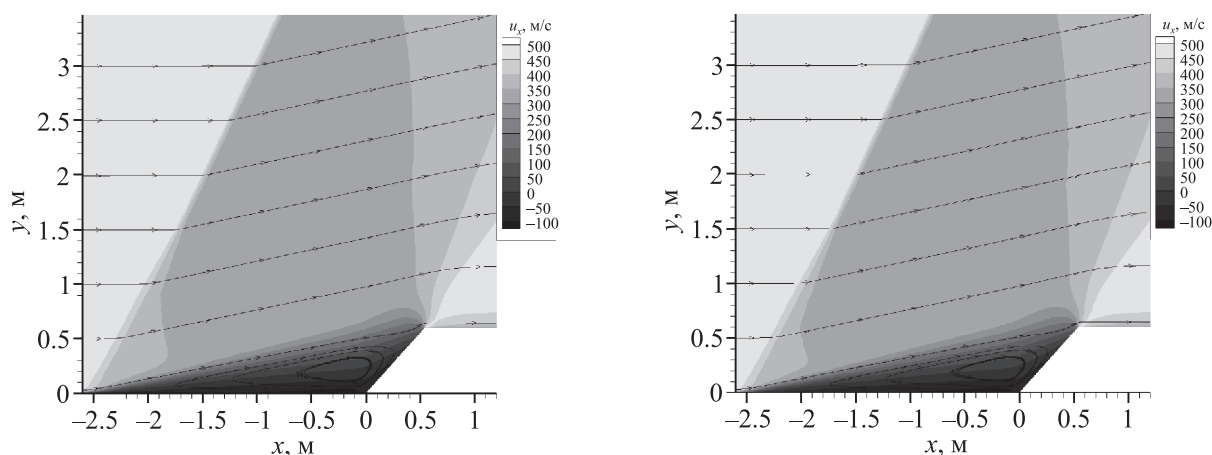


Рис. 4. Поле компоненты скорости  $v_x$  для задачи с  $\alpha = 132^\circ$ ,  $Re = 10^7$  для базовой одноблочной модели (слева) и уточняемой модели ANDD (справа)

Была проведена серия расчетов с одинаковой геометрией ( $\alpha = 132^\circ$ ), отличающихся параметрами набегающего потока ( $M = 1.5$ , число Рейнольдса  $Re$  менялось в диапазоне  $9 \cdot 10^6$ – $1.2 \cdot 10^7$ ), а также серия расчетов с одинаковыми параметрами набегающего потока, но разной геометрией обтекаемого тела (величина угла сжатия  $\alpha$  менялась в диапазоне  $128^\circ$ – $140^\circ$ ). Каждый расчет был проведен с помощью базовой модели и с помощью специальной модели (ANDD).

После этого результаты расчетов были поделены на обучающую и тестовую выборки. Расчеты обучающей выборки были использованы для обучения суррогатной модели. Обученная суррогатная модель использовалась для уточнения результатов специальной модели на тестовой выборке. Для серии расчетов с одинаковой геометрией, отличающихся числами Рейнольдса  $Re$ , было рассмотрено три сценария: интерполяция по  $Re$  и два экстраполяционных случая ( $Re_{\text{test}} < Re_{\text{train}}$  и  $Re_{\text{test}} > Re_{\text{train}}$ ). Для топологического случая (параметры набегающего потока одинаковы, варьируется величина угла  $\alpha$ ) аналогично были рассмотрены сценарии интерполяции и экстраполяции. Более подробно разбиение на обучающую и тестовую выборки для каждого сценария представлено в таблице 1.

Таблица 1. Обучающая и тестовая выборки

	Обучающая выборка		Тестовая выборка	
	$Re \cdot 10^6$	$\alpha, ^\circ$	$Re \cdot 10^6$	$\alpha, ^\circ$
1. Интерполяция по $Re$	9, 10, 12	132	11	132
2. Экстраполяция по $Re$	9, 10, 11	132	12	132
3. Экстраполяция по $Re$	10, 11, 12	132	9	132
4. Интерполяция по $\alpha$	10	128, 132, 140	10	136
5. Экстраполяция по $\alpha$	10	132, 136, 140	10	128
6. Экстраполяция по $\alpha$	10	128, 132, 136	10	140

На основании уточненных значений компонент скорости  $v_x$  и  $v_y$  для тестовой выборки численно рассчитывалось трение на стенке по формуле

$$f_\tau = \mu \left. \frac{\partial v_\tau}{\partial y} \right|_{y=0}, \quad (11)$$

где  $\mu$  — динамическая вязкость,  $v_\tau$  — касательная компонента скорости,  $y$  — расстояние по нормали к стенке.

По полученному трению рассчитывался коэффициент трения  $C_f$ :

$$C_f = \frac{F_\tau}{0.5\rho v_\infty^2 S}, \quad (12)$$

где  $F_\tau$  — интеграл от трения  $f_\tau$  по поверхности обтекаемого тела,  $v_\infty$  — скорость набегающего потока,  $\rho$  — плотность,  $S$  — площадь поверхности обтекаемого тела.

Интегрирование проводилось по части стенки, соответствующей области, где находится вихрь ( $x \in [-2.5, 0.5]$ ).

Полученные результаты для тестовой выборки сравнивались с результатами одноблочного расчета и расчета методом ANDD. На рис. 5–10 показано сравнение профилей компоненты скорости  $v_x$  вблизи стенки и трения на стенке для базовой модели (Single block), уточняемой модели (ANDD) и уточняющей суррогатной модели (RF-NDD). Для наилучшей иллюстрации результатов для всех сценариев показаны профили скорости с координатой  $x$ , выбранной из области, соответствующей наклонному участку стенки, так как именно в этой области наблюдается максимальное различие между расчетом по базовой и специальной моделям.

В таблице 2 показаны полученные коэффициенты трения  $C_f$  для базовой, уточняемой и уточненной моделей, а также относительная ошибка:

$$\varepsilon = \frac{|C_f^{sb} - C_f^{2b}|}{|C_f^{sb}|}. \quad (13)$$

Здесь  $C_f^{sb}$  — это коэффициент трения, полученный при расчете по базовой модели,  $C_f^{2b}$  — коэффициент трения, полученный при расчете по уточняемой модели или уточненной суррогатной модели соответственно.

Из таблицы 2 видно, что относительная ошибка (сравнительно с базовой моделью) при расчете коэффициента трения  $C_f^{sb}$  для суррогатной модели значительно меньше, чем для уточняемой.

Таблица 2. Коэффициенты трения  $C_f$  для базовой модели (Single Block), уточняемой модели (ANDD), уточненной модели (RF-NDD) и относительная ошибка  $\varepsilon$

	$C_f \cdot 10^{-4}$			$\varepsilon$	
	Single Block	ANDD	RF-NDD	ANDD	RF-NDD
1. Интерполяция по Re	-3.24	-2.64	-3.31	0.1849	0.0237
2. Экстраполяция по Re	-3.19	-2.6	-3.25	0.1841	0.0183
3. Экстраполяция по Re	-3.35	-2.71	-3.51	0.1891	0.0504
4. Интерполяция по $\alpha$	-3.35	-2.74	-3.38	0.1822	0.009
5. Экстраполяция по $\alpha$	-3.32	-2.71	-3.54	0.1822	0.0686
6. Экстраполяция по $\alpha$	-3.15	-2.59	-3.32	0.1757	0.0557

### Интерполяция по Re

На рис. 5 показаны результаты для сценария 1 (интерполяция по числу Рейнольдса Re). Из графиков и из таблицы 2 видно, что построенный оператор с высокой точностью уточняет результат расчета, проведенного по приближенной модели.

### Экстраполяция по Re

На рис. 6 и рис. 7 показаны результаты для сценариев 2 и 3 (экстраполяция по числу Рейнольдса Re, верхняя и нижняя границы диапазона). Эти задачи являются более сложными, чем предыдущая, поскольку число Рейнольдса для тестовой выборки находится за пределами диапазона чисел Рейнольдса для обучающей выборки. Однако результаты, полученные при уточнении модели ANDD с помощью построенного оператора, довольно близки к результатам точной модели, хоть и с меньшей точностью, чем в задаче интерполяции по Re.

### Топологический случай

На рис. 8–10 показаны результаты для сценариев 4, 5, 6 — интерполяция и экстраполяция по величине угла сжатия  $\alpha$ . Из графиков видно, что результаты получились немного хуже, чем для задачи интерполяции/экстраполяции по числу Рейнольдса Re, однако из профилей скорости, а также из таблицы 2 видно, что результаты, полученные для уточненной модели, все еще лучше, чем результаты расчетов, проведенных по приближенной модели ANDD. Также видно, что для этих случаев имеются проблемы с гладкостью решения, в связи с чем имеются небольшие участки, где графики трения для точной и уточняющей суррогатной моделей не совпадают.

В таблице 3 показано изменение интегральной характеристики (коэффициент трения  $C_f$ ) в зависимости от количества данных в обучающей выборке для задач экстраполяции по Re и  $\alpha$ . Для случаев 1 и 3 в обучающую выборку добавлялись данные, более близкие к тестовой выборке, для случаев 2 и 4 — менее близкие к тестовой выборке. Из таблицы 3 видно, что добавление в обучающую выборку данных ведет к уменьшению ошибки, причем добавление данных, более близких к тестовым случаям, значительно уменьшает ошибку, а добавление данных, менее близких к тестовым, оказывает менее значительное влияние.



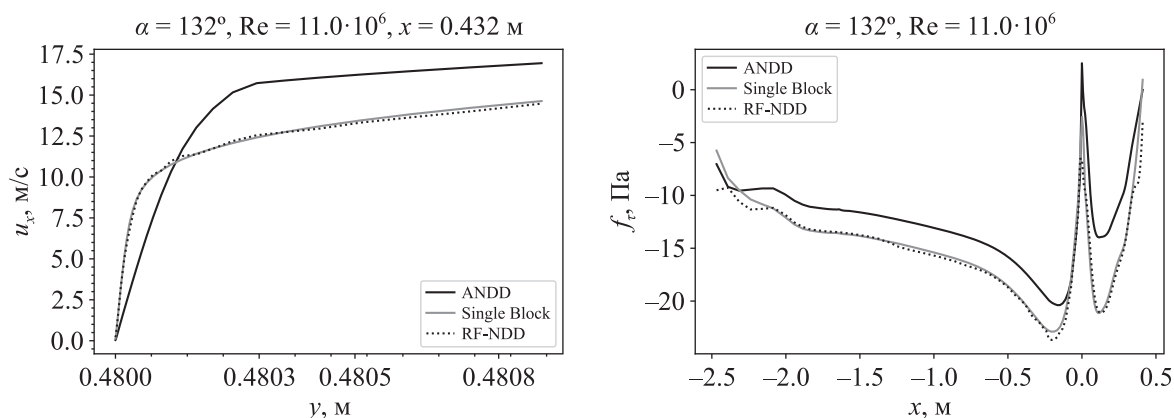


Рис. 5. Профиль компоненты скорости  $v_x$  при  $x = 0.432$  м (слева) и трение на стенке (справа) для задачи с  $\alpha = 132^\circ$ ,  $Re = 11 \cdot 10^6$  (интерполяция по  $Re$ ). Представлены результаты для базовой модели (Single Block), уточняемой модели (ANDD) и результат уточнения (RF-NDD)

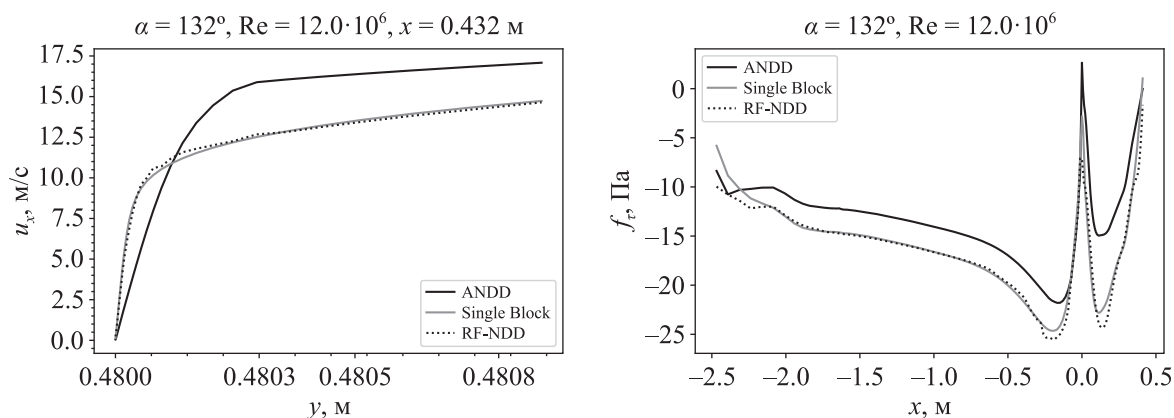


Рис. 6. Профиль компоненты скорости  $v_x$  при  $x = 0.432$  м (слева) и трение на стенке (справа) для задачи с  $\alpha = 132^\circ$ ,  $Re = 12 \cdot 10^6$  (экстраполяция по  $Re$ ). Представлены результаты для базовой модели (Single Block), уточняемой модели (ANDD) и результат уточнения (RF-NDD)

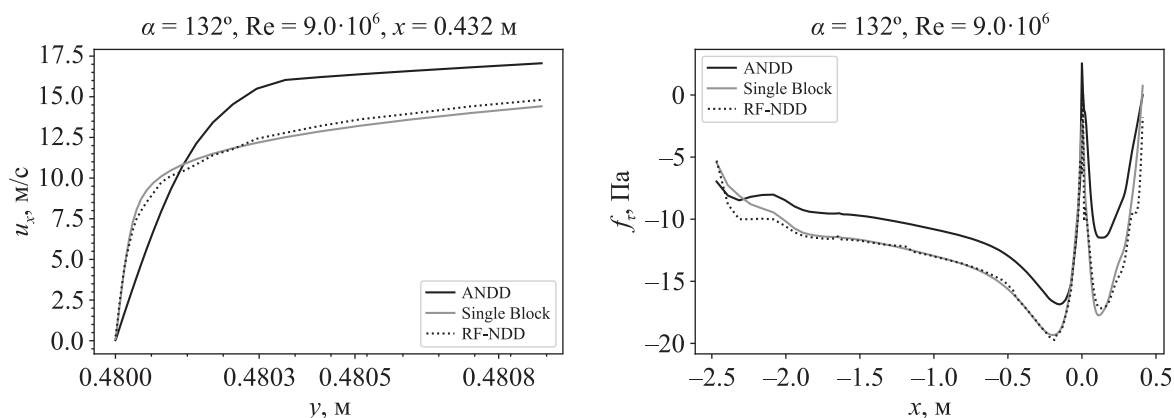


Рис. 7. Профиль компоненты скорости  $v_x$  при  $x = 0.432$  м (слева) и трение на стенке (справа) для задачи с  $\alpha = 132^\circ$ ,  $Re = 9 \cdot 10^6$  (экстраполяция по  $Re$ ). Представлены результаты для базовой модели (Single Block), уточняемой модели (ANDD) и результат уточнения (RF-NDD)

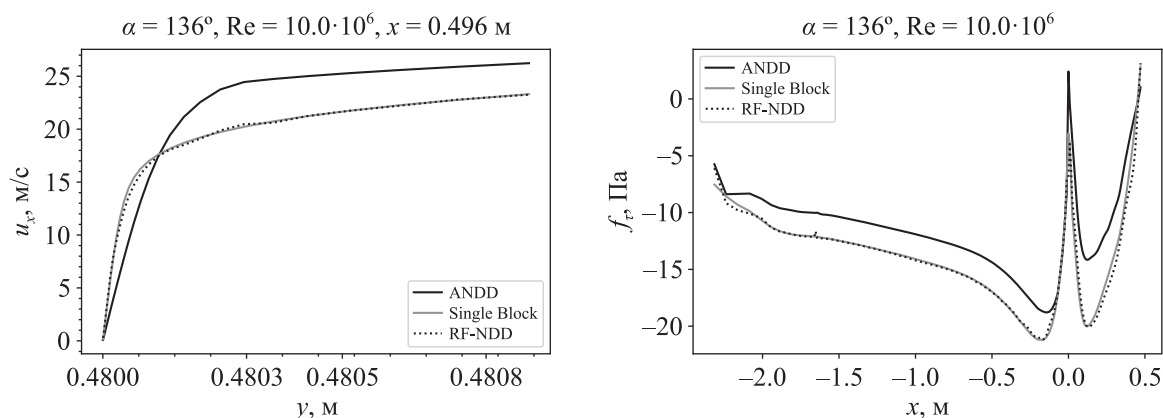


Рис. 8. Профиль компоненты скорости  $v_x$  при  $x = 0.496$  м (слева) и трение на стенке (справа) для задачи с  $\alpha = 136^\circ$ ,  $Re = 10^7$  (интерполяция по  $\alpha$ ). Представлены результаты для базовой модели (Single Block), уточняемой модели (ANDD) и результат уточнения (RF-NDD)

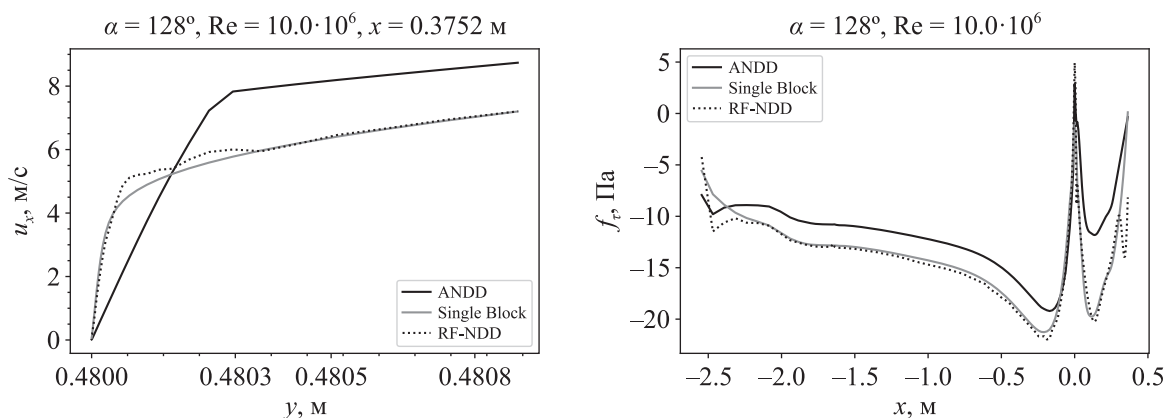


Рис. 9. Профиль компоненты скорости  $v_x$  при  $x = 0.375$  м (слева) и трение на стенке (справа) для задачи с  $\alpha = 128^\circ$ ,  $Re = 10^7$  (экстраполяция по  $\alpha$ ). Представлены результаты для базовой модели (Single Block), уточняемой модели (ANDD) и результат уточнения (RF-NDD)

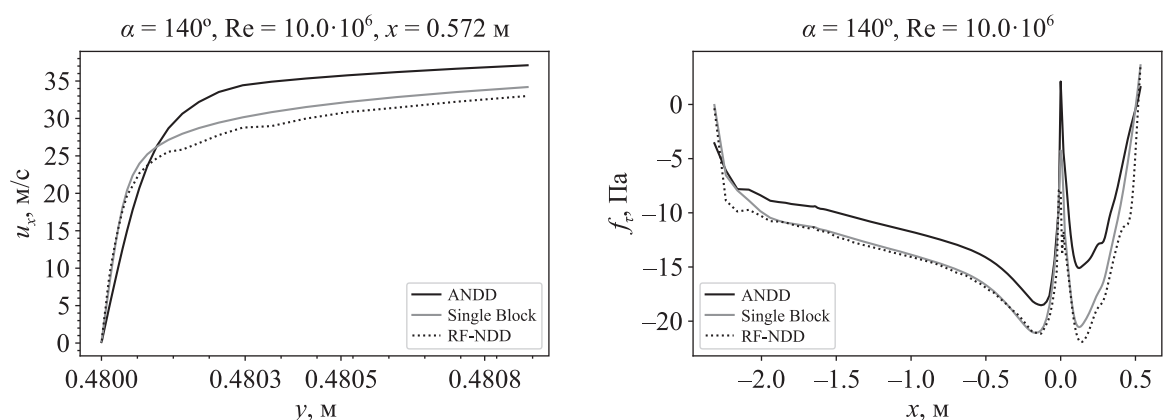


Рис. 10. Профиль компоненты скорости  $v_x$  при  $x = 0.572$  м (слева) и трение на стенке (справа) для задачи с  $\alpha = 140^\circ$ ,  $Re = 10^7$  (экстраполяция по  $\alpha$ ). Представлены результаты для базовой модели (Single Block), уточняемой модели (ANDD) и результат уточнения (RF-NDD)

Таблица 3. Зависимость коэффициента трения  $C_f$  от размера обучающей выборки

№	Тестовая выборка		Обучающая выборка		$C_f$	$\varepsilon$
	$Re \cdot 10^6$	$\alpha, \circ$	$Re \cdot 10^6$	$\alpha, \circ$		
1	12	132	11	132	-3.35	0.0506
			10, 11	132	-3.27	0.025
			9, 10, 11	132	-3.25	0.0183
2	12	132	9	132	-3.47	0.0872
			9, 10	132	-3.31	0.037
			9, 10, 11	132	-3.25	0.0183
3	10	140	10	136	-3.44	0.0928
			10	132, 136	-3.35	0.064
			10	128, 132, 136	-3.32	0.0557
4	10	140	10	128	-3.6	0.1457
			10	128, 132	-3.46	0.0985
			10	128, 132, 136	-3.32	0.0557

## Заключение

В работе был исследован способ построения оператора, уточняющего результаты расчетов, полученных с помощью менее точной (но более вычислительно эффективной) модели, на основании схожих течений, полученных более точной моделью, с использованием алгоритма Random Forest. На примере задачи сверхзвукового турбулентного обтекания угла сжатия было рассмотрено несколько сценариев: интерполяция и экстраполяция по числу Рейнольдса  $Re$ , а также топологический случай. Во всех сценариях результаты уточняющей суррогатной модели (профили скорости, а также рассчитанные на основании поля скорости трение  $f_t$  и коэффициент трения  $C_f$ ) оказались достаточно близки к результатам, полученным с помощью более точной базовой модели. Предложенный метод не только обладает высокой точностью, но также является вычислительно эффективным. Это связано с тем, что при проведении серийных расчетов число расчетов  $m$ , которые проводятся базовой и уточняемой моделью и используются для обучения суррогатной модели, достаточно мало, а остальные  $N - m$  расчетов серии проводятся быстрой специальной моделью и уточняются с помощью суррогатной модели. В то же время для топологического случая (интерполяция/экстраполяция по углу сжатия  $\alpha$ ) результаты получились хуже, чем для задач интерполяции/экстраполяции по числу Рейнольдса. Также в работе было показано, что добавление данных в обучающую выборку ведет к уменьшению ошибки.

Возможным направлением дальнейших исследований могло бы стать построение уточняющего оператора нелокальным образом, то есть учитывая влияние всего поля течения на ошибку в конкретной точке, например с помощью сверточных нейронных сетей. Это могло бы привести к лучшим результатам для топологического случая и повысить гладкость решения.

## Список литературы (References)

- Петров М. Н., Титарев В. А., Утюжников С. В., Чикиткин А. В. Многопоточная OpenMP-реализация метода LU-SGS с использованием многоуровневой декомпозиции неструктурированной расчетной сетки // Журнал вычислительной математики и математической физики. — 2017. — Т. 57, № 11. — Р. 1895–1905.
- Petrov M. N., Titarev V. A., Utyuzhnikov S. V., Chikitkin A. V. A multithreaded OpenMP implementation of the LU-SGS method using the multilevel decomposition of the unstructured computational mesh // Computational Mathematics and Mathematical Physics. — 2017. — Vol. 57, No. 11. — P. 1856–1865. — DOI: 10.1134/S0965542517110124 (Original Russian paper: Petrov M. N., Titarev V. A., Utyuzhnikov S. V., Chikitkin A. V. Mnogopotochnaja OpenMP-relizatsija metoda LU-SGS s ispolzovaniem mnogourovnevoj dekompozitsii nestrukturirovannoj raschetnoj setki // Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki. — 2017. — Vol. 57, No. 11. — P. 1895–1905.)*

- Breiman L.* Random forests // *Machine learning*. — 2001. — Vol. 45, Iss. 1. — P. 5–32. — DOI: 10.1023/A:1010933404324
- Chikitkin A. V., Petrov M. N., Titarev V. A., Utyuzhnikov S. V.* Parallel versions of implicit LU-SGS method // *Lobachevskii Journal of Mathematics*. — 2018. — Vol. 39, Iss. 4. — P. 503–512. — DOI: 10.1134/S1995080218040054
- Chikitkin A., Utyuzhnikov S., Petrov M., Titarev V.* Non-overlapping domain decomposition for modeling essentially unsteady near-wall turbulent flows // *Computers & Fluids*. — 2020. — Vol. 202. — 104506. — DOI: 10.1016/j.compfluid.2020.104506
- Craft T. J., Gant S. E., Lacovides B. E., Launder B. E.* A new wall function strategy for complex turbulence flows // *Numerical Heat Transfer. Part B: Fundamentals*. — 2004. — Vol. 45, No. 4. — P. 301–318. — DOI: 10.1080/10407790490277931
- Ferrero A., Iollo A., Larocca F.* Field inversion for data-augmented RANS modelling in turbomachinery flows // *Computers & Fluids*. — 2020. — Vol. 201. — 104474. — DOI: 10.1016/j.compfluid.2020.104474
- Grotjans H., Menter F. R.* Wall functions for industrial applications // *Computational Fluid Dynamics* — 1998. — Vol. 1, part 2. — P. 1112–1117.
- Guo X., Wei Li, Iorio F.* Convolutional neural networks for steady flow approximation // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. — 2016. — P. 481–490. — DOI: 10.1145/2939672.2939738
- Hanna B. N., Dinh N. T., Youngblood R. W., Bolotnov I. A.* Machine-learning based error prediction approach for coarse-grid Computational Fluid Dynamics (CG-CFD) // *Progress in Nuclear Energy*. — 2020. — Vol. 118. — 103140. — DOI: 10.1016/j.pnucene.2019.103140
- Kalitzin G., Medic G., Laccario G., Durbin P.* Near-wall behavior of RANS turbulence models and implications for wall functions // *Journal of Computational Physics*. — 2005. — Vol. 204, Iss. 1. — P. 265–291. — DOI: 10.1016/j.jcp.2004.10.018
- Ling J., Templeton J.* Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty // *Physics of Fluids*. — 2015. — Vol. 27, No. 8. — DOI: 10.1063/1.4927765
- Ling J., Kurzawski A., Templeton J.* Reynolds averaged turbulence modelling using deep neural networks with embedded invariance // *Journal of Fluid Mechanics*. — 2016. — Vol. 807. — P. 155–166. — DOI: 10.1017/jfm.2016.615
- Parish E. J., Duraisamy K.* A paradigm for data-driven predictive modeling using field inversion and machine learning // *Journal of Computational Physics*. — 2016. — Vol. 305. — P. 758–774. — DOI: 10.1016/j.jcp.2015.11.012
- Petrov M. N., Tambova A. A., Titarev V. A., Utyuzhnikov S. V., Chikitkin A. V.* FlowModellium Software Package for Calculating High-Speed Flows of Compressible Fluid // *Computational Mathematics and Mathematical Physics*. — 2018. — Vol. 58, No. 11. — P. 1865–1886. — DOI: 10.1134/S0965542518110118
- Petrov M., Utyuzhnikov S., Chikitkin A., Titarev V.* On extension of near-wall domain decomposition to turbulent compressible flows // *Computers & Fluids*. — 2020. — Vol. 210. — 104629. — DOI: 10.1016/j.compfluid.2020.104629

- Raissi M., Perdikaris P., Karniadakis G.E.* Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations // *Journal of Computational Physics*. — 2019. — Vol. 378. — P. 686–707. — DOI: 10.1016/j.jcp.2018.10.045
- Singh A.P., Duraisamy K.* Using field inversion to quantify functional errors in turbulence closures // *Physics of Fluids*. — 2016. — Vol. 28, No. 4. — 045110. — DOI:10.1063/1.4947045
- Singh A.P., Duraisamy K., Zhang Z.J.* Augmentation of turbulence models using field inversion and machine learning // 55th AIAA Aerospace Sciences Meeting. — Grapevine, Texas, 2017. — DOI:10.2514/6.2017-0993
- Spalart P., Allmaras S.* A one-equation turbulence model for aerodynamic flows // 30th Aerospace Sciences Meeting and Exhibit, Aerospace Sciences Meetings. — AIAA Paper 1992-0439. — P. 5–21. — DOI: 10.2514/6.1992-439
- Tracey B., Duraisamy K., Alonso J.J.* A machine learning strategy to assist turbulence model development // 53rd AIAA Aerospace Sciences Meeting. — Kissimmee, Florida, 2015. — DOI: 10.2514/6.2015-1287
- Utyuzhnikov S., Petrov M., Chikitkin A., Titarev V.* On Extension of Near-Wall Non-overlapping Domain Decomposition to Essentially Unsteady Turbulent Flows // *Proceedings of the Conference 50 Years of the Development of Grid-Characteristic Method*. — 2018. — P. 199–209.