

УДК 532.517

DOI: 10.15372/PMTF202215201

## УСОВЕРШЕНСТВОВАНИЕ RANS-МОДЕЛЕЙ С ПОМОЩЬЮ МЕТОДА СЛУЧАЙНОГО ЛЕСА С ТЕНЗОРНЫМ БАЗИСОМ ДЛЯ ТУРБУЛЕНТНЫХ ТЕЧЕНИЙ В ДВУМЕРНЫХ КАНАЛАХ С ВЫСТУПАМИ

А. Бернар, С. Н. Яковенко\*

Новосибирский национальный исследовательский государственный университет,  
Новосибирск, Россия

\* Институт теоретической и прикладной механики им. С. А. Христиановича СО РАН,  
Новосибирск, Россия

E-mails: alix.bernard9@gmail.com, s.yakovenko@mail.ru

Путем обработки результатов DNS- и RANS-расчетов течений в двумерных каналах с выступами получены входные и выходные данные для метода машинного обучения, который используется с целью усовершенствования модели анизотропии напряжений Рейнольдса и, следовательно, повышения точности RANS-подхода. В качестве технологии машинного обучения выбран метод случайного леса с тензорным базисом. Результаты, полученные с помощью новой модели для тензора анизотропии напряжений Рейнольдса, лучше согласуются с данными DNS-метода для двух геометрий течения в канале, чем результаты, полученные с помощью стандартной линейной модели вихревой вязкости.

**Ключевые слова:** моделирование турбулентности, напряжение Рейнольдса, машинное обучение, случайный лес

**Введение.** Вихреразрешающие методы (DNS, LES, измерения) позволяют получить высокоточные (HF) данные, однако требуют больших затрат. В настоящее время и в ближайшем будущем RANS-подход, при котором аппроксимируется тензор напряжений Рейнольдса, являющийся неизвестным слагаемым, будет оставаться основным инструментом описания течений жидкости и газа [1] вследствие относительно малых вычислительных ресурсов. Однако данный метод является менее точным, чем DNS- и LES-методы. Использование модифицированных RANS-моделей позволяет улучшить результаты, но их расхождение с HF-данными остается значительным [2]. Существует новый, основанный на имеющихся HF-данных для канонических течений способ разработки более точных и универсальных моделей турбулентности, предполагающий использование методов машинного обучения (ML) [3].

Технологии ML позволяют построить отображения между данными, полученными с помощью HF-метода и приближенной модели, минимизируя отклонения между ними, и найти оптимальный способ коррекций модели с учетом физических особенностей течения. Применение ML-стратегии в механике жидкости и газа можно оценивать как передовое на-

---

Работа выполнена в рамках государственного задания Института теоретической и прикладной механики СО РАН.

правление (см., например, [3, 4]). Этому способствовали проведение высокопроизводительных расчетов, позволяющих получить базы HF-данных для характеристик канонических течений, и создание алгоритмов машинного обучения для обработки больших массивов данных. В работе [4] приведены основные понятия ML-технологии, описаны методы, применяемые для построения моделей турбулентности, управления потоком и решения других актуальных задач. Методы машинного обучения могут работать в режиме черного ящика (для нейронных сетей) или позволяют получать явные корректирующие выражения для замыкания (в случае символьной регрессии) уравнений RANS- и LES-методов. ML-автоматизация калибровки RANS-аппроксимаций дает возможность, в частности, построить полные нелинейные алгебраические модели напряжений Рейнольдса для широкого класса задач [5–7].

В настоящей работе исследуется возможность с помощью методов машинного обучения усовершенствовать базовую линейную модель вихревой вязкости (LEVM) для потоков, ограниченных стенками каналов, на дне которых имеются выступы, вызывающие отрыв и повторное присоединение. В качестве ML-метода выбран метод случайного леса с тензорным базисом (TBRF). Метод построен по аналогии с нейронной сетью с тензорным базисом (TBNN) [5] и позволяет достаточно точно описать анизотропию напряжений Рейнольдса (RSA) в канонических задачах [6, 7]. Для обучения и тестирования рассматриваются течения в каналах с сужением-расширением (CD) и периодически расположенными холмами (PH) [8].

**Модели и алгоритмы расчета.** Для замыкания RANS-уравнений, описывающих несжимаемые течения, требуется аппроксимация напряжений Рейнольдса  $\tau_{ij}$ . Для определения  $\tau_{ij}$  часто используется LEVM-выражение — линейная функция тензора деформации  $S_{ij}$ :

$$\tau_{ij} = \frac{2}{3} k \delta_{ij} - 2\nu_t S_{ij}, \quad S_{ij} \equiv \frac{1}{2} \left( \frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} \right). \quad (1)$$

Турбулентная вязкость  $\nu_t = C_\mu k^2 / \varepsilon = k / \omega$  ( $C_\mu = 0,09$ ) определяется с использованием величин  $k$ ,  $\varepsilon$  или  $\omega$ , полученных с помощью моделей  $k-\varepsilon$  или  $k-\omega$ . Для улучшения модели выражение (1) можно обобщить, представив безразмерный RSA-тензор  $b_{ij}$  для несжимаемой среды в виде линейной комбинации 10 изотропных базисных тензоров  $T^{(1)}, \dots, T^{(10)}$ , зависящих от нормализованных тензоров деформации  $s_{ij}$  и вращения  $r_{ij}$  [5]:

$$b_{ij} \equiv \frac{\tau_{ij}}{2k} - \frac{1}{3} \delta_{ij} = \sum_{m=1}^{10} g^{(m)}(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) T_{ij}^{(m)}, \quad (2)$$

$$T_{ij}^{(1)} = s_{ij}, \quad T_{ij}^{(2)} = s_{ik} r_{kj} - s_{ik} r_{kj}, \quad \dots;$$

$$\lambda_1 = s_{ij} s_{ji}, \quad \lambda_2 = r_{ij} r_{ji}, \quad \lambda_3 = s_{ij} s_{jk} s_{ki}, \quad \lambda_4 = r_{ij} r_{jk} s_{ki}, \quad \lambda_5 = r_{ij} r_{jk} s_{km} s_{mi},$$

$$s_{ij} = \frac{k}{\varepsilon} S_{ij}, \quad r_{ij} = \frac{k}{\varepsilon} R_{ij}, \quad R_{ij} \equiv \frac{1}{2} \left( \frac{\partial U_i}{\partial x_j} - \frac{\partial U_j}{\partial x_i} \right).$$

ML-алгоритм заключается в поиске оптимальных функций скалярных инвариантов  $\lambda_1, \dots, \lambda_5$  для коэффициентов  $g^{(m)}$ . После определения  $g^{(m)}$  при обучении модели на HF-данных соотношение (2) используется для получения  $b_{ij}$ , встраиваемого в RANS-решатель в виде  $\tau_{ij} = 2k[b_{ij} + (1/3)\delta_{ij}]$ . Эффективность модифицированного решателя затем проверяется для течений, не участвовавших в обучении.

Для реализации дерева решений с тензорным базисом (TBDT) применяется алгоритм дерева классификации и регрессии (CART) [6, 7]. При обучении дерева решений CART пространство входных признаков многократно расщепляется на две корзины, обучающие

данные и коэффициенты  $g^{(m)}$  аппроксимируются постоянными значениями в каждой корзине. После минимизации функции стоимости решение для  $\mathbf{g}$  принимает вид

$$\mathbf{g} = (g^{(1)}, \dots, g^{(10)}) = \sum_{n=1}^{N_s} \{\hat{T}_n^T \hat{b}_n\} / \sum_{n=1}^{N_s} \{\hat{T}_n^T \hat{T}_n + \Gamma I\}, \quad (3)$$

где  $\hat{T}_n$ ,  $\hat{b}_n$  — тензоры  $T^{(m)}$  и  $b$  в уплощенной форме; индекс  $n$  соответствует элементу обучающей выборки (случайно выбранному из набора данных);  $N_s$  — количество выборок;  $I$  — единичная матрица;  $\Gamma$  — задаваемый малым параметр регуляризации. В [6, 7] уплощение по индексам  $(i, j)$  в (2) в каждой  $n$ -й точке проводилось в виде  $\hat{b}_n = [(b_n)_{11}, (b_n)_{12}, \dots, (b_n)_{23}, (b_n)_{33}]$ , т. е. для записи матриц  $3 \times 3 \times n$  для  $b$  и  $3 \times 3 \times n \times t$  для  $T^{(m)}$  в более компактном виде  $9 \times n$  и  $9 \times n \times t$ . В методе TBRF используется несколько TBDT-деревьев с пакетированием. Окончательным результатом, полученным с помощью TBRF, является средняя или медианная величина для всех деревьев решений. Подробно алгоритмы CART, TBDT, TBRF описаны в [6, 7].

**Параметры и результаты моделирования.** Для проверки ML-подхода выбраны два течения в двумерных каналах (рис. 1). Первое течение (CD), выбранное и для обучения, и для тестирования, наблюдается в канале с сужением-расширением и с развитым турбулентным входным потоком при  $Re_0 = U_0 h / \nu = 12\,600$  ( $h$  — половина высоты канала;  $U_m$  — максимальная входная скорость). Второе течение (PH) наблюдается в канале, на дне которого периодически расположены холмы, при  $Re = U_b H / \nu = 5600$  ( $H$  — высота холма;  $U_b$  — среднерасходная скорость на входе). Для обоих случаев доступны данные, полученные для различных уклонов  $\alpha$  холма с помощью DNS-метода и по  $(k-\omega)$ -модели с

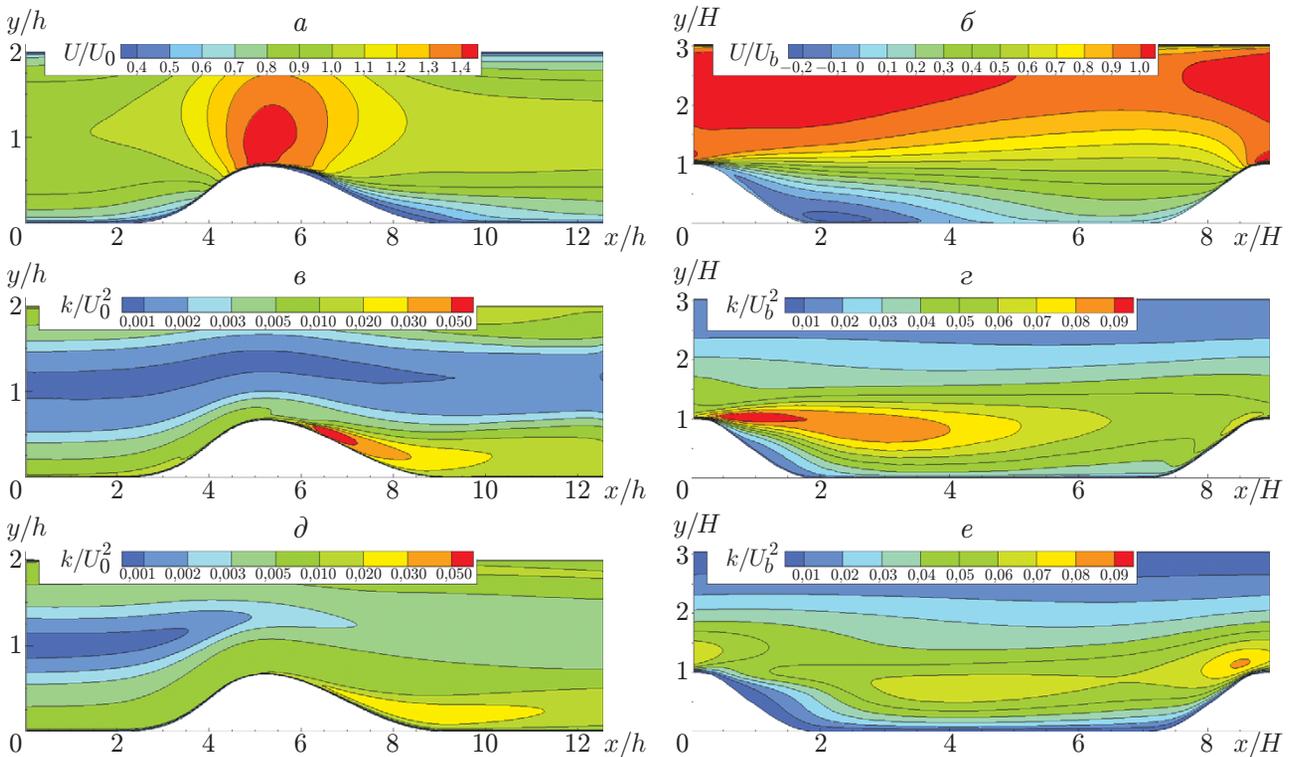


Рис. 1. Распределения продольной скорости (а, б) и турбулентной кинетической энергии (в–е), полученные в DNS-расчетах (а–г) и RANS-расчетах (д, е) [8]: а, в, д — течение CD, б, г, е — течение PH при  $\alpha = 1$

LEVM-выражением [8]. При обучении выбираем течение РН при  $\alpha = 1,2$ , а для априорной проверки уточненных с помощью TBRF-аппроксимации компонент тензора  $b_{ij}$  — течение РН при  $\alpha = 1,0$ , не участвовавшее в обучении.

Для моделирования RSA с использованием ML-алгоритмов большое значение имеет задание входных признаков, которые представляют различные аспекты динамики жидкости и газа. В отличие от  $\lambda_1, \dots, \lambda_5$  в TBNN коэффициенты  $g^{(m)}$  в TBRF зависят от большего количества признаков  $N_f = 17$  [7]. В настоящей работе  $N_f = 15$ , два признака, содержащие  $\tau_{ij}$ , не используются (с этим изменением результаты получаются близкими). Базисные тензоры и входные признаки берутся из результатов, полученных с помощью RANS-модели, а  $b_{ij}$  — из эталонных HF-данных [8].

Для ML-алгоритма, реализованного средствами Python, высокоуровневого объектно-ориентированного интерпретируемого языка программирования общего назначения, выбраны следующие гиперпараметры. Число TBDT в TBRF задано равным  $N_T = 20$  согласно работе [6], в которой показано, что определенная в [5] путем суммирования отдельных компонент  $b_{ij}$  и точек данных  $N_d$  среднеквадратичная ошибка (RMSE)

$$e = \sqrt{\frac{1}{6N_d} \sum_{i=1}^3 \sum_{j=1}^i \sum_{n=1}^{N_d} (b_{ij,\text{TBRF}} - b_{ij,\text{DNS}})^2} = \sqrt{\frac{1}{6} \sum_{i=1}^3 \sum_{j=1}^i e_{ij}^2},$$

$$e_{ij}^2 = \frac{1}{N_d} \sum_{n=1}^{N_d} (b_{ij,\text{TBRF}} - b_{ij,\text{DNS}})^2 \quad (4)$$

уменьшается с ростом  $N_T$  как в случае потока в квадратном канале, так и в случае течения РН, а при  $N_T > 20$  поведение  $e$  становится более устойчивым, шум снижается. При уменьшении  $N_d$  до 6000 получаем репрезентативный набор данных для течений РН и CD (по 3000 точек на каждое течение), достаточный для приемлемой работы ML-алгоритма. Число выборок, используемых в загрузочном наборе данных (данных, выбранных случайным образом для обучения каждого TBDT), задается равным  $N_s = 0,4N_d$ . Количество признаков, выбираемых при расщеплении, устанавливается равным  $N_{f,s} = 5$ , т. е. приблизительно в три раза меньше, чем  $N_f$ , как и при использовании стандартного метода случайного леса. Для течения РН число  $N_{f,s} = 11$  дает меньшую RMSE, однако это улучшение  $e$  по сравнению со случаем  $N_{f,s} = 5$  незначительно [6]. По аналогии минимальное количество выборок в узлах листа TBDT задано равным  $N_{s,l} = 1$ , как принято по умолчанию для методов случайного леса, поскольку значение  $e$  в этом случае несущественно отличается от  $e$  при оптимальном числе  $N_{s,l} = 9$  [6]. Заметим, что небольшие значения  $N_s, N_{f,s}, N_{s,l}$  выбраны для экономии вычислительных ресурсов. Параметр  $\Gamma$  в (3) задан равным 0,1, так как использование значения  $\Gamma < 0,1$  не позволяет достичь видимого улучшения.

Далее выполняем алгоритм TBRF. На рис. 2 показано, что модель (2) с TBRF-модификацией позволяет получить более точные данные для значений  $b_{ij}$ , чем базовая RANS-модель (1). Это также показано в таблице, где RMSE для RANS-результатов находится из (4) путем подстановки  $b_{ij,\text{LEVM}}$  вместо  $b_{ij,\text{TBRF}}$ .

**Заключение.** В работе метод TBRF применен для уточнения LEVM-модели (1) с использованием высокоточных данных DNS в качестве эталонного решения для канонических течений в каналах с выступами на дне. ML-алгоритм реализован с более полным использованием свойств языка Python, что позволяет получать результаты на 10–40 % быстрее (для одного TBDT), чем с помощью доступных кодов [6, 7]. Оценки показывают, что значение тензора анизотропии напряжения Рейнольдса, найденное с помощью TBRF,

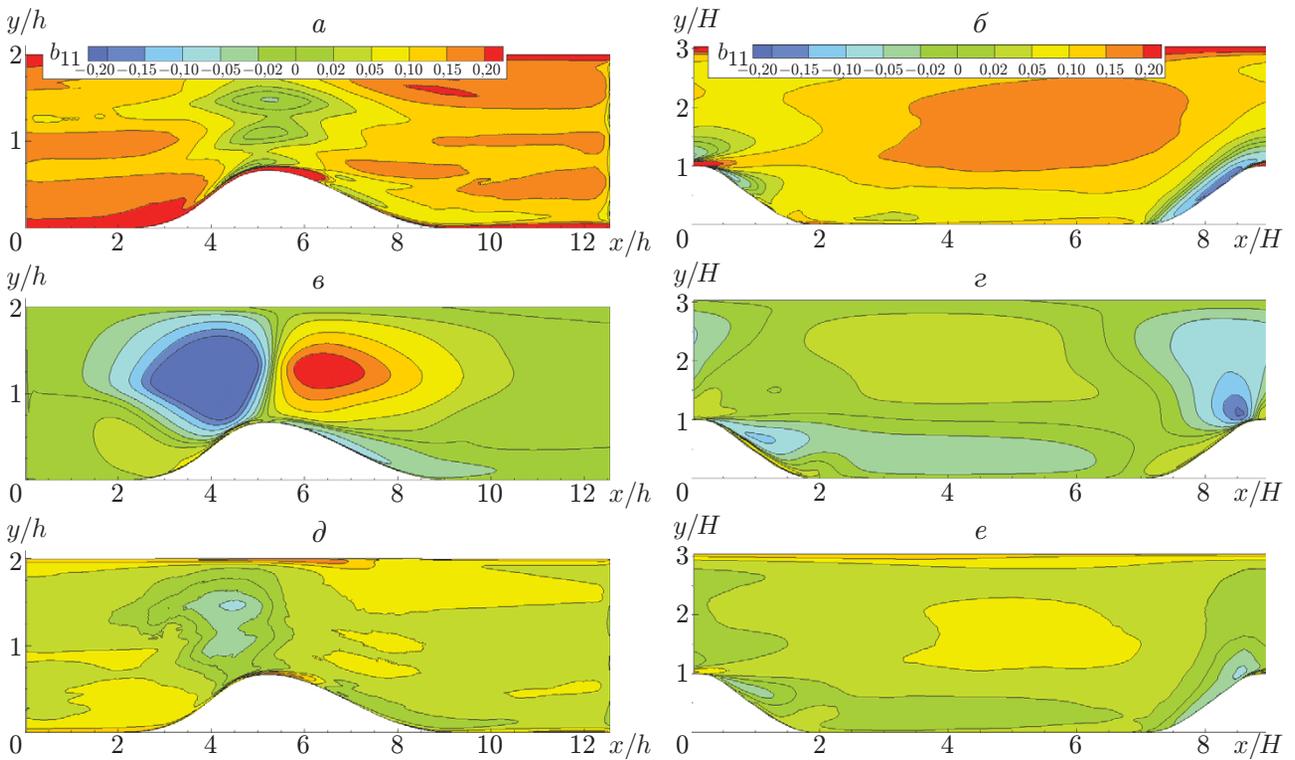


Рис. 2. Распределения  $b_{11}$ , полученные с использованием различных методов:  $a, б$  — DNS-расчет [8],  $в, г$  — RANS-модель [8],  $д, е$  — TBTRF-модель (настоящая работа);  $a, в, д$  — течение CD,  $б, г, е$  — течение PH при  $\alpha = 1$

Среднеквадратические ошибки для компонент RSA-тензора  $b_{ij}$

Тип течения	$e_{11}$		$e_{22}$		$e_{33}$		$e_{12}$		$e$	
	LEVM	TBRF	LEVM	TBRF	LEVM	TBRF	LEVM	TBRF	LEVM	TBRF
CD	0,277	0,193	0,212	0,140	0,098	0,073	0,090	0,063	0,152	0,105
PH	0,178	0,128	0,169	0,122	0,101	0,075	0,059	0,057	0,111	0,082

ближе к эталонным решениям, чем аналогичное значение, полученное с помощью базовой RANS-модели.

В дальнейшем целесообразно проверить, применимы ли алгоритмы машинного обучения и их параметры, рассмотренные в работе, для других течений и можно ли при встраивании в CFD-решатели RSA-аппроксимаций, построенных с помощью ML-алгоритмов, получить более точные результаты. Также к усовершенствованию TBRF может привести реализация автоматического выбора признаков при расщеплении в каждом узле каждого TBDT-дерева.

### ЛИТЕРАТУРА

- Durbin P.** Some recent developments in turbulence closure modeling // Annual Rev. Fluid Mech. 2018. V. 50. P. 77–103.
- Yakovenko S. N., Chang K. C.** Computational studies of near-wall behaviors of low-Reynolds-number Reynolds-stress models // AIAA J. 2019. V. 51. P. 279–296.

3. **Duraisamy K., Iaccarino G., Xiao H.** Turbulence modeling in the age of data // Annual Rev. Fluid Mech. 2019. V. 51. P. 357–377.
4. **Brunton S. L., Noack B. R., Koumoutsakos P.** Machine learning for fluid mechanics // Annual Rev. Fluid Mech. 2020. V. 52. P. 477–508.
5. **Ling J., Kurzawski A., Templeton J.** Reynolds averaged turbulence modelling using deep neural networks with embedded invariance // J. Fluid Mech. 2016. V. 807. P. 155–166.
6. **Kaandorp M.** Machine learning for data-driven RANS turbulence modelling: Master thesis. Delft: Delft Univ. of Technol., 2018.
7. **Kaandorp M., Dwight R. P.** Data-driven modelling of the Reynolds stress tensor using random forests with invariance // Comput. Fluids. 2020. V. 202. 104497.
8. **McConkey R., Yee E., Lien F. S.** A curated dataset for data-driven turbulence modelling // Scientific Data. 2021. V. 8. 255.

*Поступила в редакцию 1/IX 2022 г.,  
после доработки — 10/X 2022 г.  
Принята к публикации 27/X 2022 г.*

---