

УЧИТЫВАНИЕ ПРОМЫШЛЕННОГО ЗАГРЯЗНЕНИЯ В МОДЕЛЯХ ДЫХАНИЯ ПОЧВЫ

Сморкалов И.А.¹, Бедин Д.А.²

¹*Институт экологии растений и животных УрО РАН, г. Екатеринбург, Россия*

ivan.a.smorkalov@gmail.com

²*Институт математики и механики УрО РАН, г. Екатеринбург, Россия*

bedin@imm.uran.ru

Аннотация: Учет уровня загрязнения улучшил качество аппроксимации дыхания почвы в районах действия металлургических предприятий. Приближение методами машинного обучения оказалось лучше аппроксимации модифицированной базовой моделью, но найденные зависимости имеют неявный характер и хуже интерпретируются.

1. Введение

Необходимость моделирования глобального цикла углерода уже ни у кого не вызывает сомнения. Для улучшения качества моделей важно определить ключевые факторы, влияющие на интенсивность основных компонентов цикла, одним из которых является дыхание почвы.

Вопросы математического моделирования дыхания почвы разрабатываются уже более 50 лет. В качестве предикторов, в основном, используют температуру почвы и/или её влажность. Однако модели, использующие только эти параметры, работают в очень ограниченных рамках масштаба и времени и не учитывают других значимых средовых факторов. Например, промышленное загрязнение. Ранее было установлено, что сильное загрязнение тяжелыми металлами может как снижать, так и почти не влиять на уровень почвенной эмиссии CO₂ (Ramsey et al., 2005; Kozlov et al., 2009; Копчик et al., 2015; Сморкалов, Воробейчик, 2016) даже при значительной дигрессии экосистем.

Цель работы – оценить качество аппроксимации дыхания почвы в районах действия металлургических предприятий с учетом фактора загрязнения и без него.

2. Объекты и методы

Использовали данные по дыханию, температуре, влажности почвы и воздуха в районах действия двух крупных металлургических предприятий: Среднеуральского и Карабашского медеплавильных заводов (СУМЗ и КМЗ, соответственно). В районе СУМЗ (подзона южной тайги) работы проведены в ельниках-пихтарниках, в районе КМЗ (предлесостепные сосново-березовые леса) – в производных березняках, образовавшихся на месте сосновых лесов. В обоих районах хорошо выражены фоновая (в районе СУМЗ на удалении 20–33 км от завода, КМЗ – 18–32 км), буферная (СУМЗ – 4–10 км, КМЗ – 9–18 км) и импактная (СУМЗ – 1–3 км, КМЗ – 1–5 км) зоны, характеризующие последовательные стадии техногенной дигрессии лесных экосистем.

Измерения были проведены в 2011–2013 гг. и в 2018 г. в разгар вегетационных сезонов (конец июня – начало августа) и по их завершении (конец сентября – конец октября) на 30 пробных площадях у каждого из заводов (по 3 площади на 10 удалениях от 1 до 33 км от источника загрязнения). Всего использовано 4110 измерений. Дыхание почвы определяли закрытым динамическим камерным методом полевыми респирометрами SR1LP (Qubit Systems, Канада) и Li-8100A (Li-Cor biosciences, США), температуру на глубине 5 см – термометром из комплекта респирометров, объемную влажность почвы – датчиком ThetaProbe ML2x (Delta-T Devices, Великобритания). В качестве оценки уровня загрязнения использовали кислото-растворимую концентрацию меди в лесной подстилке (Сморкалов, Воробейчик, 2016).

Полную выборку измерений разделили на тестовую (30%) и обучающую (70%) выборки со стратификацией по пробным площадям. Были обработаны пропуски в измерениях температуры и влажности.

В качестве базовой модели (БМ) была принята модель, введённая в работе (Алферов и др., 2017). В БМ скорость дыхания имеет экспоненциальную зависимость от температуры почвы и зависимость от влажности типа обратной параболы. Коэффициенты БМ подбирали нелинейным методом наименьших квадратов, используя метод оптимизации Левенберга – Марквардта. Вычисления проводили при помощи пакета научных вычислений scikit-learn (Pedregosa *et al.*, 2011) для языка Python.

Была предпринята попытка улучшить приближение данных при помощи современных методов машинного обучения. В качестве признаков использовали температуру и влажность минеральных горизонтов, температуру, влажность и запас подстилки, температуру воздуха, концентрацию меди. Регрессию строили при помощи случайного леса («Random forest», Hastie *et al.*, 2009). Параметры леса (количество деревьев, максимальное количество признаков, рассматриваемых для разделения, минимальное количество данных на листе) подобрали на обучающей выборке поиском по сетке, качество оценивали путём кросс-валидации с 5 фолдами. После чего, лес с наилучшими параметрами дообучали на полной обучающей выборке.

3. Результаты

При применении БМ на тестовой выборке был показан коэффициент $R^2 = 0.2$, среднеквадратичное отклонение (СКО) между предсказанием и измерением равнялось 2.8 (в обучающей выборке значения характеристик приближения существенно не отличались). Для проверки БМ была применена к отдельной выборке в районе КМЗ для фонового уровня загрязнения. Значение коэффициента детерминации на тестовой подвыборке равнялось 0.32.

Для моделирования влияния загрязнения ввели дополнительный множитель в БМ, зависящий от концентрации загрязняющего вещества по типу сигмоиды: при малых концентрациях множитель близок к единице и имеет плато, при больших концентрациях почти равен нулю, наибольшая скорость изменения приходится на некоторую пороговую концентрацию. Добавление такого члена позволило улучшить приближение: на тесте был показан коэффициент $R^2 = 0.39$ и СКО = 2.5 (на обучении 0.36 и 2.5, соответственно). Улучшение по сравнению с БМ подтверждается изменением критерия Акаике с 7.91 до 6.36.

Однако, ещё более хорошее приближение было показано с использованием случайного леса. Так, для модели, использующей влажность почвы, её температуру и уровень Cu , на тестовой выборке было показано качество $R^2 = 0.65$ и СКО = 1.8. При дополнительном учёте воздушной температуры качество возросло: $R^2 = 0.74$ и СКО = 1.5 (на обучающей выборке 0.96 и 0.7, соответственно).

Таким образом, учёт загрязнения улучшил качество приближения данных. По показателю «важности» («feature importances») при использовании случайного леса выделили четыре переменные: температура воздуха, температура почвы, «загрязнение», влажность почвы, которые следует использовать для попытки перевода результатов машинного обучения в формульную зависимость.

Работа выполнена при финансовой поддержке комплексной программы УрО РАН №18-4-4-44.

Литература

- Алферов А.М., Блинов В.Г., Гитарский М.Л. и др. Мониторинг потоков парниковых газов в природных экосистемах. / Под ред. Д.Г. Замолотчикова, Д.В. Карелина, М.Л. Гитарского, В.Г. Блинова. Саратов: Амирит, 2017. 279 с.
- Копчик Г.Н., Кадулин М.С., Захарова А.И. Влияние техногенного загрязнения на эмиссию диоксида углерода почвами в Кольской Субарктике // Журнал общей биологии. 2015. Т. 76. № 1. С. 48-62.

- Сморкалов И.А., Воробейчик Е.Л. Механизм стабильности эмиссии CO₂ из лесной подстилки в условиях промышленного загрязнения // Лесоведение. 2016. № 1. С. 34-43.
- Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. New York: Springer-Verlag, 2009. 745 p.
- Kozlov M.V., Zvereva E.L., Zverev V.E. Impacts of point polluters on terrestrial biota: Comparative analysis of 18 contaminated areas. Dordrecht: Springer, 2009. 466 p.
- Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research. 2011. V. 12, P. 2825-2830.
- Ramsey P.W., Rillig M.C., Feris K.P. et al. Relationship between communities and processes: new insights from a field study of a contaminated ecosystem // Ecology Letters. 2005. V. 8. № 11. P. 1201-1210.