

ĐÁNH GIÁ SỰ TƯƠNG QUAN GIỮA NỒNG ĐỘ CADIMI VÀ TÍNH CHẤT ĐẤT, THỜI TIẾT TẠI TỈNH BÌNH DƯƠNG

TRẦN THỊ ANH THU¹, NGUYỄN LÊ TẤN ĐẠT¹,
VONGPHACHANH SOUKNILANH¹

¹Khoa Tài nguyên môi trường, Trường Luật và Quản lý phát triển,
Trường Đại học Thủ Dầu Một

Tóm tắt:

Nghiên cứu này nhằm mục đích đánh giá sự tương quan giữa nồng độ Cadimi (Cd) và tính chất đất, thời tiết tại tỉnh Bình Dương. Nghiên cứu thu thập dữ liệu (kết quả quan trắc Cd trong đất, tính chất đất, yếu tố thời tiết), phân tích tương quan, thống kê. Kết quả nghiên cứu cho thấy, dữ liệu thu thập nồng độ Cd trong đất tại Bình Dương giai đoạn 2021-2022 có sự biến đổi theo thời gian và không gian; tuy nhiên, không có mối tương quan rõ ràng giữa nồng độ Cd và các yếu tố thời tiết như nhiệt độ, độ ẩm và lượng mưa. Mối quan hệ giữa nồng độ Cd với các yếu tố trong đất như độ pH, độ ẩm và thành phần dinh dưỡng cho thấy có ít tác động đến sự tích tụ Cd trong đất. Phân tích sai số trong mô hình hồi quy tuyến tính nhấn mạnh sự cần thiết phải nâng cao độ chính xác và tin cậy của kết quả, đồng thời, việc tối ưu hóa khả năng dự đoán nồng độ Cd không cần phụ thuộc vào các yếu tố thời tiết.

Từ khóa: Cadimi, tính chất đất, thời tiết, tương quan, hồi quy tuyến tính.

Ngày nhận bài: 2/12/2024; Ngày sửa chữa: 30/12/2024; Ngày duyệt đăng: 5/1/2025.

Evaluation of the correlation between Cadmium concentration and soil properties, weather in Binh Duong province

Abstract:

This study aims to evaluate the correlation between Cadmium (Cd) concentration and soil properties, weather in Binh Duong province. The study collects data (Cd monitoring results in soil, soil properties, weather factors), analyzes correlations, and statistics. The study results show that the collected data on Cd concentration in soil in Binh Duong in 2022 varies over time and space; however, there is no clear correlation between Cd concentration and weather factors such as temperature, humidity, and rainfall. The relationship between Cd concentration and soil factors such as pH, moisture, and nutrient content shows that there is little significant impact on Cd accumulation in soil. Error analysis in the linear regression model emphasizes the need to improve the accuracy and reliability of the results, and shows that optimizing the ability to predict Cd concentration does not depend on weather factors.

Keywords: Cadmium, Soil Properties, Weather, Correlation, Linear Regression.

JEL Classifications: N53, N57, R52.

1. GIỚI THIỆU

Tỉnh Bình Dương nằm trong vùng kinh tế trọng điểm của Đông Nam bộ, với sự phát triển mạnh mẽ trong các ngành công nghiệp, nông nghiệp, làng nghề và khai khoáng. Việc sử dụng các loại hóa chất trong nông nghiệp đã dẫn đến nguy cơ ô nhiễm kim loại nặng trong đất. Ngoài ra, các hoạt động sản xuất công nghiệp, khai thác khoáng sản như than đá, quặng chì, quặng thiếc cũng góp phần làm tăng mức ô nhiễm kim loại nặng trong môi trường đất và nước, với các chất độc hại như Asen (As), Chì (Pb), Cadimi (Cd), Crom (Cr).

Cd là một kim loại độc hại, chất gây ô nhiễm có tính di động cao, dễ dàng xâm nhập vào chuỗi thức ăn và đe dọa sức khỏe con người cũng như môi trường sinh thái. Cd xâm nhập vào đất thông qua các

hoạt động của con người, tích lũy trong cây trồng gây rối loạn các quá trình sinh hóa và sinh lý. Các ion Cd_2^+ được hấp thụ bởi rễ cây và vận chuyển đến các bộ phận ở trên của cây, gây ra những tác động bất lợi đến sự phát triển của cây trồng, dẫn đến đe dọa sức khỏe con người.

Mặc dù, theo kết quả quan trắc của Trung tâm Quan trắc Kỹ thuật - Tài nguyên và môi trường Bình Dương nồng độ Cd trong đất chưa vượt mức ô nhiễm tuy nhiên, cần đánh giá tình hình gia tăng nồng độ kim loại nặng để phòng ngừa và ứng biến kịp thời các tác động xấu đến môi trường và sức khỏe con người. Do vậy, với xu thế phát triển theo hướng công nghiệp hóa mạnh mẽ, trong tương lai việc giải quyết vấn đề ô nhiễm Cd trong đất là một thử thách lớn, đặc biệt đối với khu vực khai thác khoáng sản, làng



nghe sơn mài, ngành công nghiệp hóa chất, sản xuất nông nghiệp lạm dụng phân bón hóa học. Kỹ thuật, công nghệ xử lý, làm sạch đất khó hơn gấp nhiều lần làm sạch không khí, sạch nước, bởi vì kim loại nặng khó tách khỏi thổ nhưỡng, khác hẳn với các chất thải hữu cơ. Nơi bị ô nhiễm nhẹ cũng phải mất 3 - 5 năm vẫn chưa thể khôi phục được hoàn toàn như cũ.

Ngoài ra, các nghiên cứu thông thường tiến hành đo đạc nồng độ Cd trong đất theo cách truyền thống là đến vị trí quan trắc thực hiện việc lấy mẫu đất, tiến hành phân tích nồng độ Cd trong mẫu đất. Với cách quan trắc mẫu như vậy sẽ gây mất nhiều thời gian, công sức và chi phí thực hiện nên các nguồn dữ liệu thực đo còn rất hạn chế. Nguồn dữ liệu quan trắc này cũng chỉ có thể cung cấp thông tin trong khoảng thời gian ngắn tại vị trí cố định, khó có thể sử dụng để phân tích tổng thể toàn tỉnh Bình Dương nhằm theo dõi biến động trong khoảng thời gian dài. Do vậy, rất cần thiết cho việc phân tích tương quan giữa nồng độ Cd với các yếu tố về tính chất đất, thời tiết khu vực để cho thấy khả năng dự báo sự biến thiên nồng độ Cd theo thời gian.

Nghiên cứu này chủ yếu tập đánh giá sự tương quan giữa nồng độ Cd và tính chất đất, thời tiết tại tỉnh Bình Dương. Mục tiêu cụ thể của nghiên cứu này bao gồm: 1) Phân tích diễn biến các mẫu Cd giai đoạn 2021-2022; 2) Tương quan giữa Cd và các yếu tố trong đất và yếu tố thời tiết.

2. DỮ LIỆU VÀ PHƯƠNG PHÁP

2.1. Thu thập dữ liệu

2.1.1. Dữ liệu quan trắc Cd

Nồng độ Cd trong đất tại tỉnh Bình Dương được quan trắc định kì 6 tháng/lần (tháng 4 và tháng 10) (TTQTBD, 2022). Lượng mẫu đối với mỗi điểm quan trắc được lấy đầy đủ số lượng, ngoài ra theo Thông tư số 10/2021/TT-BTNMT ngày 30/06/2021 của Bộ TN&MT quy định kỹ thuật quan trắc môi trường và quản lý thông tin, dữ liệu quan trắc chất lượng môi trường, tổng số mẫu quan trắc và mẫu QA/QC là 26 mẫu/đợt theo quyết định phê duyệt mạng lưới quan trắc môi trường của tỉnh Bình Dương.

Dữ liệu vị trí quan trắc Cd, bao gồm thông tin về các mẫu đất thu thập từ nhiều khu vực khác nhau, phân chia thành hai loại đất chính: Đất xám và đất đỏ vàng. Trong số 18 mẫu, các mẫu đất công nghiệp được ký hiệu từ ĐCN1 đến ĐCN11, với vị trí lấy mẫu chủ yếu tập trung tại các khu công nghiệp lớn như KCN Đồng An, KCN Việt Hương II và KCN Sóng Thần. Những khu vực này thường gắn liền với sự phát triển công nghiệp và đặc điểm đất tại đây chủ yếu là đất xám và đất đỏ vàng.

Ngoài ra, các mẫu đất đô thị, ký hiệu từ ĐĐT1 đến ĐĐT7, được thu thập từ các khu vực như trung tâm thị xã Dĩ An, thị xã Thuận An và thành phố mới Bình Dương. Đất ở những khu vực này chủ yếu là đất đỏ vàng và đất xám, phản ánh sự đa dạng trong điều kiện đất đai của vùng đô thị.

2.1.2. Dữ liệu khí hậu

Dữ liệu thời tiết toàn cầu từ NASA/POWER, bao gồm các biến như nhiệt độ ở độ cao 2 mét (T2M), điểm sương (T2MDEW), nhiệt độ bầu ướt (T2MWET), nhiệt độ bề mặt đất (TS) và lượng mưa đã hiệu chỉnh (PRECTOTCORR), được coi là dữ liệu thứ cấp phục vụ cho nghiên cứu. Nguồn dữ liệu này cung cấp cái nhìn tổng quan về điều kiện khí hậu và thời tiết, giúp các nhà nghiên cứu phân tích xu hướng khí hậu, dự báo thời tiết và đánh giá tác động của các yếu tố khí hậu lên môi trường. Dưới đây là bảng dữ liệu mô tả các biến thời tiết toàn cầu từ dữ liệu NASA-POWER, cùng với thông tin về nguồn gốc và ý nghĩa của từng biến.

Bảng 1. Dữ liệu khí hậu

TT	Biến	Mô tả	Đơn vị
1	T2M	Nhiệt độ ở độ cao 2 mét	°C
2	T2MDEW	Điểm sương/băng giá ở độ cao 2 mét	°C
3	T2MWET	Nhiệt độ không khí ướt ở độ cao 2 mét	°C
4	TS	Nhiệt độ bề mặt đất	°C
5	T2M_RANGE	Biên độ nhiệt độ MERRA-2 ở độ cao 2 mét	°C
6	T2M_MAX	Nhiệt độ tối đa MERRA-2 ở độ cao 2 mét	°C
7	T2M_MIN	Nhiệt độ tối thiểu MERRA-2 ở độ cao 2 mét	°C
8	QV2M	Độ ẩm riêng ở độ cao 2 mét	g/kg
9	RH2M	Độ ẩm tương đối ở độ cao 2 mét	%
10	PRECTOTCORR	Lượng mưa đã hiệu chỉnh MERRA-2	mm/ngày

2.1.3. Dữ liệu tính chất đất

Các mẫu đất thu thập từ hai loại khu vực: Đất công nghiệp và đất đô thị. Các mẫu đất được lấy từ nhiều vị trí khác nhau, phản ánh tính chất của chúng. Đối với đất công nghiệp, các chỉ số như tỉ trọng, dung trọng, độ ẩm và pH trong dung dịch KCl và nước cho thấy đặc tính vật lý và hóa học của đất tại các khu công nghiệp như KCN Đồng An và KCN Việt Hương II. Tương tự, đất đô thị cũng có các chỉ số này, với mẫu được thu thập từ các khu vực như trung tâm thị xã Dĩ An và trung tâm thành phố mới Bình Dương. Tỉ trọng và dung trọng cho biết độ nén và

khả năng giữ nước, trong khi độ ẩm phản ánh lượng nước có trong đất, rất quan trọng cho sự sinh trưởng của thực vật. Độ pH, đo lường độ axit hoặc kiềm, ảnh hưởng đến khả năng hấp thụ chất dinh dưỡng của cây trồng.

2.2. Phân tích ma trận tương quan

Ma trận tương quan là bảng chứa các hệ số tương quan giữa các biến trong nghiên cứu, với mỗi phần tử đại diện cho mối quan hệ giữa hai biến cụ thể. Hệ số tương quan nằm trong khoảng từ -1 đến 1:

Tương quan dương ($0 < r \leq 1$): Khi một biến tăng, biến kia cũng tăng. Ví dụ, hệ số tương quan 0.75 giữa lượng mưa và năng suất cây trồng cho thấy mối liên hệ tích cực.

Tương quan âm ($-1 \leq r < 0$): Khi một biến tăng, biến kia giảm. Ví dụ, hệ số -0.60 giữa độ pH và nồng độ Cd cho thấy, mối liên hệ tiêu cực.

Không tương quan (r gần 0): Không có mối quan hệ rõ ràng. Ví dụ, hệ số tương quan gần 0 giữa chiều cao cây và nhiệt độ cho thấy không có liên hệ đáng kể.

Cách tính Hệ số tương quan. Hệ số tương quan thường được tính bằng phương pháp Pearson, được định nghĩa như sau:

$$r = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Trong đó:

- $Cov(X, Y)$ là hiệp phương sai giữa hai biến X và Y, đo lường mức độ mà hai biến thay đổi cùng nhau.
- σ_X là độ lệch chuẩn của biến X, đo lường mức độ phân tán của các giá trị xung quanh trung bình.
- σ_Y là độ lệch chuẩn của biến Y.

Ma trận tương quan giúp phát hiện mối quan hệ giữa các biến, xác định yếu tố ảnh hưởng đến kết quả nghiên cứu và cải thiện mô hình dự đoán. Nó cũng giảm thiểu đa cộng tuyến trong phân tích hồi quy, nâng cao độ tin cậy của kết quả. Ngoài ra, các tổ chức sử dụng ma trận này để ra quyết định dựa trên mối quan hệ giữa các chỉ số kinh tế - xã hội và môi trường.

2.3. Phân tích ANOVA

ANOVA (Phân tích phương sai) là một phương pháp thống kê quan trọng, được phát triển bởi Ronald A. Fisher vào những năm 1920, nhằm so sánh sự khác biệt giữa nhiều nhóm dữ liệu. Mục tiêu chính của ANOVA là kiểm tra giả thuyết về việc các nhóm có cùng giá trị trung bình hay không, điều này rất cần thiết trong các lĩnh vực như y học, sinh học, kinh tế và khoa học - xã hội. ANOVA phân tách biến thiên trong dữ liệu thành hai nguồn:

- **Biến thiên giữa các nhóm:** Do sự khác biệt giữa các nhóm.

- **Biến thiên trong nhóm:** Do sự khác biệt giữa các quan sát trong cùng một nhóm.

Phân tích ANOVA giúp xác định sự khác biệt có ý nghĩa giữa nhiều nhóm và cung cấp cái nhìn sâu sắc về các yếu tố ảnh hưởng đến biến số nghiên cứu, là công cụ không thể thiếu trong nhiều lĩnh vực nghiên cứu. Hiểu rõ lý thuyết và ứng dụng của ANOVA giúp các nhà nghiên cứu đưa ra quyết định chính xác hơn.

2.4. Phân tích thống kê trên R

Phân tích thống kê trong R là quá trình sử dụng ngôn ngữ lập trình R để thực hiện các phép toán thống kê, kiểm tra giả thuyết và phân tích dữ liệu. Dưới đây là hướng dẫn tổng quan:

1. **Cài đặt:** Người dùng cần cài đặt R và RStudio để tạo môi trường làm việc.

2. **Nhập dữ liệu:** Dữ liệu có thể được tích hợp từ nhiều nguồn như file CSV hoặc Excel.

3. **Khám phá dữ liệu:** Sử dụng các hàm như `str()` và `summary()` để hiểu cấu trúc và đặc điểm của dữ liệu, bao gồm kiểm tra giá trị thiếu.

4. **Phân tích thống kê:** Thực hiện các phép toán cơ bản như tính trung bình, độ lệch chuẩn và kiểm định t-test. R cũng cho phép tính toán ma trận tương quan để xác định mối quan hệ giữa các biến.

R cung cấp nhiều hàm và gói mở rộng, giúp người dùng thực hiện phân tích một cách hiệu quả, từ đó đưa ra quyết định dựa trên dữ liệu một cách chính xác.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Sự phân bố hàm lượng Cd trong đất

Biểu đồ phân bố hàm lượng Cd trong đất là một dạng biểu đồ boxplot (biểu đồ hộp), được sử dụng để mô tả phân phối của một biến số (trong trường hợp này là giá trị Cd) theo từng nhóm (loại đất: đất công nghiệp và đất đô thị). Dưới đây là phân tích chi tiết về các thành phần của biểu đồ này: Đất đô thị có hàm lượng Cd ổn định hơn đất công nghiệp (Hình 1).

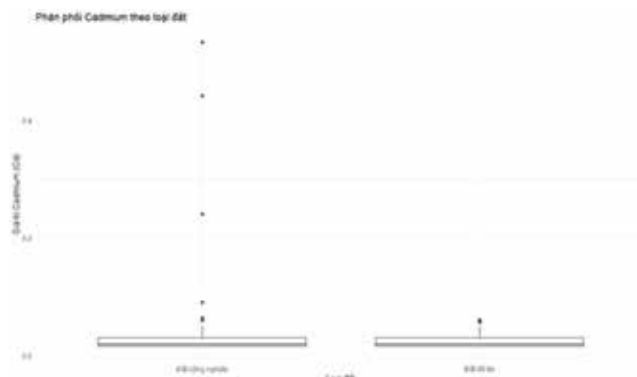
- **Trục tung (Y-axis):** Đại diện cho giá trị Cd

- **Trục hoành (X-axis):** Phân loại đất, bao gồm đất công nghiệp và đất đô thị.

Đất công nghiệp cho thấy hàm lượng Cd đa dạng và có nhiều điểm ngoại lai. Giá trị Cd trong đất công nghiệp trải dài từ 0,015 mg/kg đến 0,441 mg/kg, cho thấy sự không đồng nhất trong mức độ ô nhiễm. Các điểm ngoại lai, đặc biệt là những giá trị cao, có thể chỉ ra sự hiện diện của các hoạt động sản xuất hoặc xử lý chất thải không đúng cách. Những điểm này cần được điều tra kỹ lưỡng để xác định nguồn gốc ô nhiễm, như từ các nhà máy sản xuất hóa chất, chế

biến kim loại hoặc các sự cố tràn chất thải. Sự hiện diện của Cd ở đất công nghiệp có thể ảnh hưởng tiêu cực đến sức khỏe của con người và môi trường. Cd là một kim loại nặng, có thể gây ra tổn thương thận, ảnh hưởng đến hệ miễn dịch và có thể dẫn đến ung thư nếu tiếp xúc lâu dài. Do đó, việc giám sát chất lượng đất công nghiệp là rất quan trọng, đồng thời cần có các biện pháp khắc phục kịp thời để giảm thiểu rủi ro ô nhiễm.

Ngược lại, đất đô thị có hàm lượng Cd ổn định và chủ yếu nằm trong khoảng thấp, thường dưới 0,1 mg/kg. Điều này cho thấy, các hoạt động đô thị có xu hướng ít gây ô nhiễm Cd hơn so với đất công nghiệp. Không có nhiều điểm ngoại lai trong đất đô thị, cho thấy một sự đồng nhất trong các mẫu đất và khả năng quản lý đất tốt hơn.

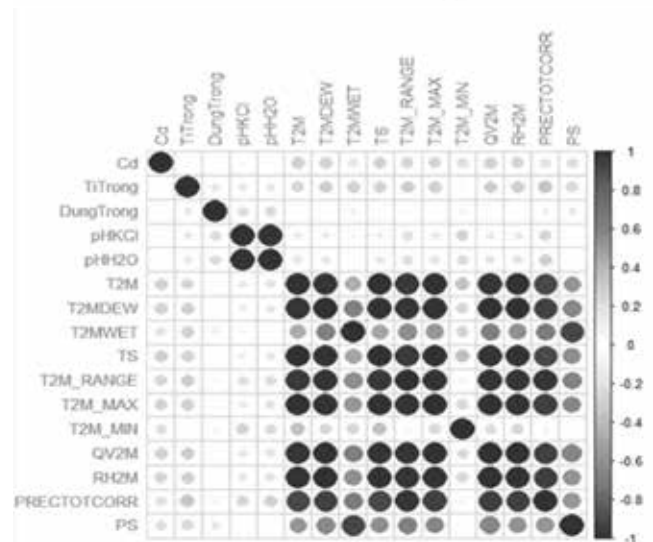
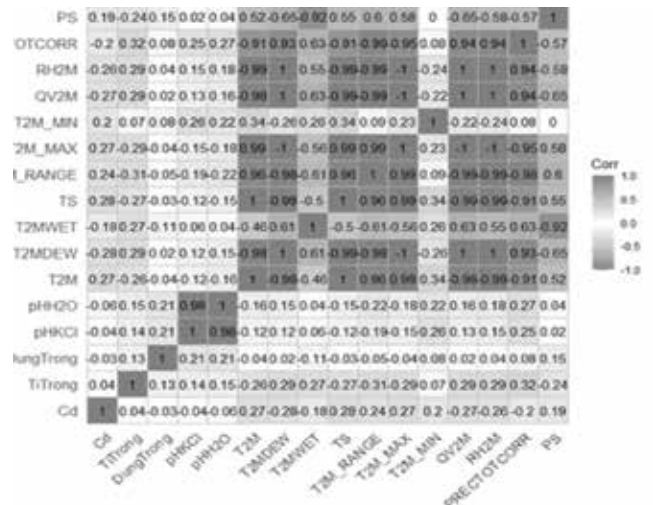


▲ Hình 1. Biểu đồ Phân phối hàm lượng Cd theo loại đất

3.2. Tương quan giữa Cd và các yếu tố trong đất

Ma trận tương quan là công cụ hữu ích trong việc phân ánh mối quan hệ giữa nồng độ Cd và các yếu tố đất, khí tượng khác (Hình 2). Trong ma trận này, các mối tương quan được thể hiện thông qua một thang màu, giúp dễ dàng nhận diện mức độ và hướng của từng mối liên hệ. Màu đỏ trong ma trận biểu thị mối tương quan dương, cho thấy rằng khi một yếu tố tăng lên, nồng độ Cd cũng có xu hướng tăng theo. Ngược lại, màu xanh dương biểu thị mối tương quan âm, chỉ ra rằng khi một yếu tố gia tăng, nồng độ Cd lại có xu hướng giảm. Sự phân loại rõ ràng này không chỉ giúp nhận diện các yếu tố có ảnh hưởng đến Cd mà còn cung cấp cơ sở cho các nghiên cứu tiếp theo nhằm hiểu rõ hơn về tác động của các yếu tố môi trường đến mức độ ô nhiễm.

Mỗi tương quan dương mạnh giữa Cd và các yếu tố liên quan đến tính chất vật lý của đất được thể hiện qua các chỉ số cho thấy, các yếu tố này có thể ảnh hưởng đến sự phân bố của Cd trong môi trường.



▲ Hình 2. Ma trận tương quan giữa nồng độ Cd và các yếu tố tính chất đất và khí hậu

- **Tỉ trọng** có giá trị tương quan là 0,039. Mặc dù con số này không quá cao, nhưng nó vẫn cho thấy một xu hướng dương giữa tỉ trọng của đất và nồng độ Cd. Điều này có thể chỉ ra rằng các loại đất có tỉ trọng cao hơn có khả năng giữ lại Cd tốt hơn. Tỉ trọng cao thường liên quan đến cấu trúc đất chặt chẽ, có thể làm giảm khả năng di chuyển của các chất ô nhiễm như Cd, từ đó dẫn đến nồng độ cao hơn trong những loại đất này. Điều này cũng có thể phản ánh rằng các loại đất này có thể chứa nhiều chất hữu cơ, giúp giữ lại Cd trong cấu trúc của chúng.

- **Về dung trọng** giá trị tương quan là -0,030. Mặc dù đây là một giá trị âm, nó vẫn cho thấy rằng có một số yếu tố liên quan đến dung trọng có thể ảnh hưởng đến sự phân bố của Cd trong đất. Dung trọng thấp thường liên quan đến đất có độ xốp cao hơn, có khả năng giữ nước và chất dinh dưỡng tốt hơn, nhưng cũng có thể dẫn đến việc Cd được rửa trôi dễ dàng hơn trong điều kiện độ ẩm cao.

Dù các giá trị tương quan này không mạnh mẽ, nhưng chúng vẫn chỉ ra rằng Cd có thể có mối liên hệ



tích cực với các đặc tính vật lý của đất. Sự tương quan này mở ra hướng nghiên cứu sâu hơn về cách mà các yếu tố vật lý của đất ảnh hưởng đến sự tích tụ và phân bố của Cd trong môi trường.

Mối tương quan âm giữa Cd và độ pH đất thể hiện rõ qua các chỉ số pH cụ thể.

- *pHKCl*: Giá trị tương quan là -0,042. Mặc dù đây là một giá trị nhỏ, nhưng nó cho thấy một sự tương quan âm giữa pHKCl và nồng độ Cd. Điều này có nghĩa là khi độ pHKCl tăng lên, nồng độ Cd có xu hướng giảm. Sự tăng cường độ pH có thể làm cho Cd ít hòa tan hơn trong đất, từ đó giảm khả năng hấp thụ của cây trồng và sự tích tụ trong hệ sinh thái đất.

- *pHH₂O*: Giá trị tương quan là -0,060, cho thấy mối liên hệ âm mạnh hơn một chút. Khi độ pHH₂O tăng, nồng độ Cd cũng có xu hướng giảm theo. Điều này có thể giải thích rằng độ pH giúp cải thiện khả năng giữ lại các ion Cd trong đất, đồng thời có thể làm giảm mức độ di chuyển của Cd trong môi trường nước. Các giá trị tương quan âm này nhấn mạnh rằng khi độ pH đất tăng, nồng độ Cd có khả năng giảm.

Điều này có thể có ý nghĩa quan trọng trong việc quản lý ô nhiễm Cd. Việc cải thiện độ pH của đất thông qua các biện pháp như bổ sung vôi có thể giúp giảm thiểu sự tích tụ của Cd, làm cho đất trở nên an toàn hơn cho cây trồng và sức khỏe con người.

Mối tương quan giữa Cd và các yếu tố khí hậu cho thấy một số liên hệ thú vị, mặc dù các giá trị tương quan ở mức độ trung bình, cho thấy rằng mối liên hệ này không quá chặt chẽ.

- *Nhiệt độ (T2M)*: Giá trị tương quan là 0,270, cho thấy mối tương quan dương. Điều này nghĩa là khi nhiệt độ tăng lên, nồng độ Cd cũng có xu hướng tăng theo. Nhiệt độ cao có thể thúc đẩy quá trình phân hủy các chất hữu cơ trong đất, dẫn đến việc giải phóng Cd từ các nguồn ô nhiễm. Sự tăng cường nhiệt độ cũng có thể ảnh hưởng đến khả năng hòa tan của Cd trong nước, làm cho nồng độ của nó trong đất tăng lên.

- *Độ ẩm (QV2M, RH2M)*: Các giá trị tương quan dao động từ khoảng -0,260 đến -0,270, cho thấy mối tương quan âm. Điều này chỉ ra rằng khi độ ẩm không khí và độ ẩm trong đất tăng lên, nồng độ Cd có xu hướng giảm. Độ ẩm cao có thể tạo điều kiện cho quá trình rửa trôi, làm giảm sự tích tụ của Cd trong đất. Tuy nhiên, sự rửa trôi này cũng có thể dẫn đến việc Cd bị di chuyển vào nguồn nước, gây ra vấn đề ô nhiễm nước.

- *Lượng mưa (PRECTOTCORR)*: Giá trị tương quan là -0,200, cho thấy mối tương quan âm. Khi lượng mưa tăng, nồng độ Cd có thể giảm do tác động của dòng chảy bề mặt và rửa trôi, giúp loại bỏ Cd khỏi đất. Tuy nhiên, lượng mưa lớn cũng có thể dẫn đến sự di chuyển của Cd vào nguồn nước, gây ra các vấn đề ô nhiễm khác.

Nhìn chung, ma trận tương quan cho thấy Cd có mối liên hệ rõ ràng với các đặc tính vật lý của đất và độ pH. Cụ thể, sự tương quan dương với tỉ trọng và mối tương quan âm với độ pH chỉ ra rằng các yếu tố này có thể ảnh hưởng đáng kể đến nồng độ Cd trong môi trường đất. Điều này gợi ý rằng việc điều chỉnh các đặc tính vật lý của đất, chẳng hạn như tỉ trọng và độ pH, có thể giúp quản lý ô nhiễm Cd hiệu quả hơn.

Tuy nhiên, mối liên hệ giữa Cd và các yếu tố khí hậu lại chỉ ở mức độ trung bình. Mặc dù có sự tương quan với nhiệt độ và độ ẩm, các giá trị không mạnh cho thấy rằng yếu tố khí hậu có thể có ảnh hưởng nhưng không phải là yếu tố quyết định duy nhất trong việc phân bố Cd. Những thông tin này sẽ rất hữu ích trong việc hiểu sâu hơn về hành vi và phân bố của Cd trong hệ thống đất. Việc nắm rõ các mối tương quan này không chỉ giúp trong việc phát triển các biện pháp quản lý ô nhiễm mà còn cung cấp cơ sở cho các nghiên cứu tiếp theo nhằm tìm hiểu cách mà Cd tương tác với các yếu tố môi trường khác. Điều này có thể dẫn đến việc xây dựng các chiến lược canh tác bền vững và cải thiện chất lượng đất, từ đó bảo vệ sức khỏe cộng đồng và môi trường.

3.3. Mô hình hồi quy và phân tích phương sai ANOVA

Nghiên cứu đã áp dụng mô hình hồi quy tuyến tính để phân tích các yếu tố ảnh hưởng đến nồng độ Cd. Mục tiêu là xác định mức độ giải thích của mô hình đối với sự biến động của nồng độ Cd và đánh giá tính phù hợp của các biến độc lập trong mô hình.

- *Hệ số xác định (R-squared)*: Giá trị R-squared là 0,09673, trong khi R-squared hiệu chỉnh (Adjusted R-squared) là 0,0283. Điều này cho thấy mô hình chỉ giải thích khoảng 2,8% biến động của nồng độ Cd, trong khi 97,2% còn lại được giải thích bởi các yếu tố khác không có trong mô hình.

- *Kiểm định F*: Kết quả kiểm định F cho thấy mô hình không có ý nghĩa thống kê với p-value = 0,2309, lớn hơn mức ý nghĩa 0,05, cho thấy mô hình không phù hợp để dự đoán nồng độ Cd.

- *Hệ số hồi quy*: Các biến như tỉ trọng, dung trọng, pHKCl và PRECTOTCORR không có ý nghĩa thống kê (p-value > 0,05). Biến T2M (nhiệt độ ở độ cao 2m) có p-value = 0,0768, gần mức ý nghĩa 0,05, cho thấy, có thể có ảnh hưởng đến nồng độ Cd, mặc dù không rõ ràng.

Tổng quan, mô hình hồi quy tuyến tính này không thực sự phù hợp và không giải thích được sự biến động của nồng độ Cd một cách hiệu quả. Nguyên nhân có thể là do chưa đưa vào đủ các biến quan trọng. Từ những kết quả này, nhận thấy rằng mô hình hồi quy không thực sự phù hợp để cải thiện khả năng giải thích và dự đoán sự biến động của nồng độ Cd.



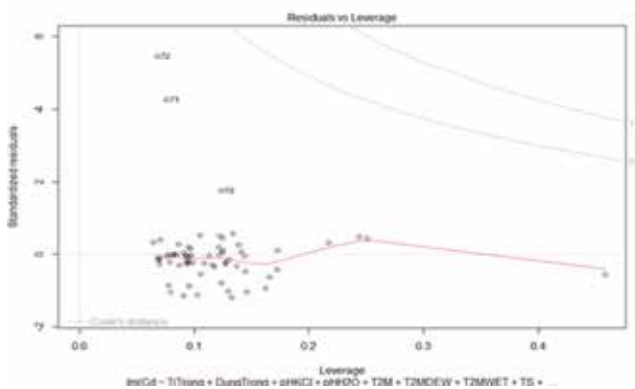
Phân tích ANOVA (Phân tích phương sai) là một công cụ thống kê quan trọng, cho phép chúng ta đánh giá sự khác biệt giữa nhiều nhóm dữ liệu. Trong nghiên cứu này, nhóm tác giả đã thực hiện phân tích ANOVA để kiểm tra nồng độ Cd giữa các nhóm khác nhau.

- **Bậc tự do (Df):** Có 71 bậc tự do cho biến "Group", tương ứng với số lượng nhóm khác nhau trong nghiên cứu. Điều này cho thấy, sự đa dạng trong các nhóm dữ liệu đang phân tích.

- **Tổng bình phương (Sum Sq):** Giá trị 0,4762 cho "Group" cho thấy có sự khác biệt đáng kể về tổng bình phương giữa các nhóm. Điều này chỉ ra rằng có sự biến thiên trong nồng độ Cd giữa các nhóm.

- **Giá trị trung bình bình phương (Mean Sq):** Giá trị 0,006707 là ước lượng phương sai của Cd giữa các nhóm. Giá trị này cho thấy mức độ phân tán của nồng độ Cd trong từng nhóm.

Từ những thông tin này, có thể kết luận, tồn tại sự khác biệt có thể có ý nghĩa giữa các nhóm về nồng độ Cd.



▲ Hình 3. Mô hình hồi quy và phân tích phương sai ANOVA

3.4. Đánh giá giả định của phương trình hồi quy

Kết quả từ bài kiểm tra Shapiro-Wilk cho sai số của mô hình cho thấy, giá trị thống kê W là 0.82298 với p-value rất nhỏ (7.522e-08), cho thấy, có khả năng cao phân phối của sai số không tuân theo phân phối chuẩn. Dựa trên các biểu đồ chẩn đoán, có thể rút ra những nhận xét sau:

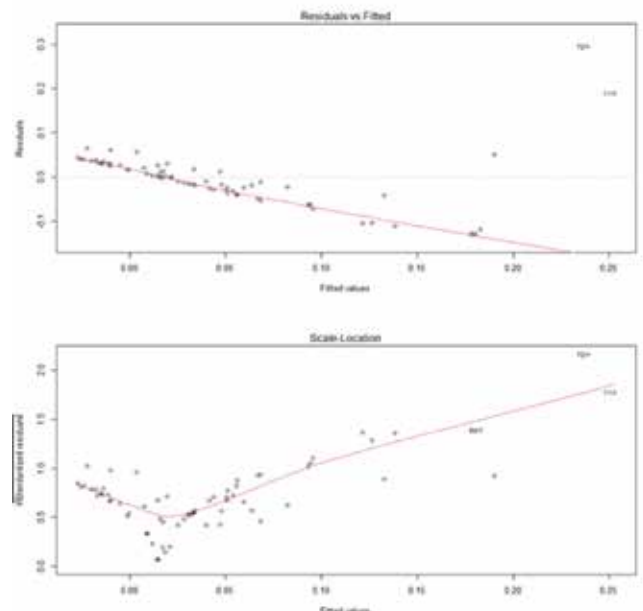
- **Dữ liệu so với Giá trị ước lượng (Residuals vs Fitted):** Biểu đồ này cho thấy mối quan hệ giữa các giá trị dữ liệu và giá trị ước lượng của mô hình. Mô hình hồi quy không thể hiện được mối quan hệ tuyến tính hoàn hảo và có một số điểm dữ liệu nằm ngoài khu vực dự đoán, gợi ý rằng mô hình có thể chưa hoàn hảo.

- **Sai số chuẩn hóa (Q-Q Residuals):** Biểu đồ này so sánh các phân vị của dữ liệu với các phân vị lý thuyết của phân phối chuẩn. Đường cong của residuals không hoàn toàn khớp với đường chuẩn, cho thấy residuals có thể không phân phối chuẩn.

- **Mức độ ảnh hưởng (Scale-Location):** Biểu đồ này thể hiện mối quan hệ giữa căn bậc hai của trị tuyệt đối của dữ liệu và các giá trị ước lượng. Sự phân tán của các điểm dữ liệu không hoàn toàn đều, gợi ý rằng giả định phương sai không đổi có thể bị vi phạm.

- **Dữ liệu so với mức độ ảnh hưởng (Residuals vs Leverage):** Biểu đồ này trình bày mối quan hệ giữa các giá trị dữ liệu và mức độ ảnh hưởng của từng quan sát. Có một số điểm dữ liệu có ảnh hưởng lớn đến mô hình, cần được xem xét kỹ lưỡng.

Tóm lại, các biểu đồ chẩn đoán cho thấy mô hình hiện tại có thể vi phạm một số giả định cơ bản, bao gồm phân phối chuẩn và phương sai không đổi của residuals. Do đó, cần xem xét cải thiện mô hình để đáp ứng tốt hơn các giả định này, nhằm đảm bảo tính chính xác và độ tin cậy của kết quả phân tích.



▲ Hình 4. Đánh giá giả định của phương trình hồi quy

3.5. Đánh giá mối liên hệ giữa Cd và các yếu tố khí hậu

Dựa trên các biểu đồ phân tán được cung cấp, nhóm tác giả có thể chia sẻ một số nhận xét về mối quan hệ giữa biến phụ thuộc Cd và các biến độc lập:

- **Cd so với Nhiệt độ tại 2 mét, Điểm sương tại 2 mét, Nhiệt độ bề mặt đất:** Các biểu đồ cho thấy không có rõ ràng mối quan hệ tuyến tính giữa Cd và các biến độc lập như Nhiệt độ tại 2 mét, Điểm sương tại 2 mét và Nhiệt độ bề mặt đất. Sự phân tán của các điểm dữ liệu cho thấy không có xu hướng nhất quán nào trong mối liên hệ giữa Cd và các biến này. Điều này có thể cho thấy rằng các yếu tố khác, ngoài nhiệt độ, có thể ảnh hưởng đến nồng độ Cd trong môi trường.

- **Cd so với Biên độ nhiệt độ tại 2 mét, Nhiệt độ tối đa tại 2 mét, Nhiệt độ tối thiểu tại 2 mét:** Tương tự,

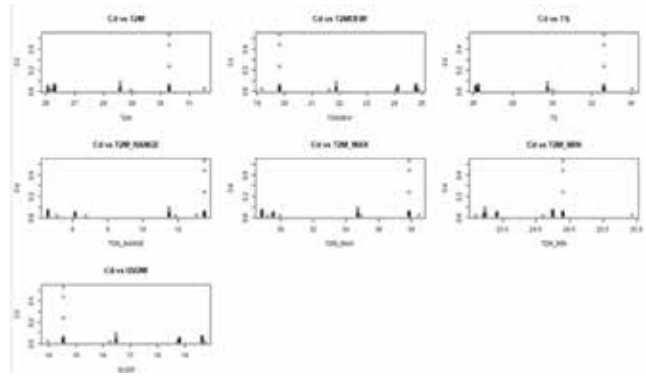
biểu đồ phân tán giữa Cd và các biến Biên độ nhiệt độ tại 2 mét, Nhiệt độ tối đa tại 2 mét và Nhiệt độ tối thiểu tại 2 mét cũng không cho thấy mối quan hệ tuyến tính rõ ràng. Các điểm dữ liệu không tạo thành một mẫu hình nhất quán, điều này càng củng cố nhận định rằng mô hình hồi quy tuyến tính có thể không phù hợp để mô tả mối quan hệ này.

- *Cd so với Độ ẩm riêng tại 2 mét*: Đáng chú ý, trong biểu đồ phân tán giữa Cd và Độ ẩm riêng tại 2 mét, có dấu hiệu của một mối quan hệ tuyến tính âm. Một số điểm dữ liệu tập trung ở phần dưới của biểu đồ cho thấy rằng khi giá trị của Độ ẩm riêng tại 2 mét tăng lên, nồng độ Cd có xu hướng giảm xuống. Mối quan hệ này có thể chỉ ra rằng độ ẩm có thể ảnh hưởng đến sự di chuyển và tích lũy của Cd trong môi trường, mặc dù cần nhiều phân tích hơn để xác nhận điều này.

Nhìn chung, các biểu đồ phân tán cho thấy khó khăn trong việc xác định mối quan hệ tuyến tính rõ ràng giữa biến phụ thuộc Cd và các biến độc lập đã đề cập. Điều này gợi ý rằng mô hình hồi quy tuyến tính có thể không phù hợp để mô tả mối quan hệ này. Việc áp dụng các mô hình phi tuyến hoặc các kỹ thuật học máy có thể giúp tìm ra các mẫu hình phức tạp hơn, từ đó nâng cao khả năng dự đoán nồng độ Cd.

Khi làm việc với dữ liệu thời tiết toàn cầu từ các nguồn như NASA/POWER, CERES hoặc MERRA2, có một số lý do chính khiến số liệu về thời tiết có thể chưa được chi tiết. Đầu tiên, dữ liệu từ các nguồn này thường có độ phân giải không gian và thời gian nhất định. Ví dụ, dữ liệu có thể được thu thập theo lưới địa lý lớn, điều này dẫn đến việc thông tin không đủ chi tiết cho các khu vực nhỏ hoặc các hiện tượng thời tiết cục bộ. Khi dữ liệu không phản ánh chính xác các điều kiện địa phương, khả năng phân tích mối quan hệ giữa các biến và nồng độ Cd sẽ bị ảnh hưởng, làm giảm độ chính xác của các kết quả nghiên cứu. Thứ hai, nhiều dữ liệu thời tiết được ước lượng từ các mô hình dự báo hoặc cảm biến vệ tinh. Các mô hình này sử dụng các phương pháp ước lượng để suy diễn các giá trị từ các điểm dữ liệu thưa thớt hơn và điều này có thể dẫn đến sai số và không chính xác đối với các giá trị cụ thể. Các yếu tố địa phương như độ cao, loại đất và các điều kiện môi trường đặc thù có thể không được mô hình hóa đầy đủ, do đó không phản ánh chính xác hiện trạng thời tiết thực tế.

Hơn nữa, sự thay đổi của thời tiết theo thời gian và không gian cũng có thể làm cho các giá trị ước lượng trở nên khó khăn trong việc theo dõi và phân tích. Những yếu tố này kết hợp lại có thể gây ra những thách thức trong việc sử dụng dữ liệu thời tiết để rút ra các kết luận chính xác về mối quan hệ giữa các biến môi trường và nồng độ Cd.



▲ Hình 5. Hình biểu đồ phân tán giữa Cd và các yếu tố khí hậu

4. KẾT LUẬN VÀ KIẾN NGHỊ

Nghiên cứu này đáp ứng được tính cấp thiết trong việc giảm ô nhiễm kim loại nặng trong đất. Kết quả nghiên cứu cho thấy dữ liệu thu thập nồng độ Cd trong đất tại Bình Dương giai đoạn 2021-2022 có sự biến đổi theo thời gian và không gian; tuy nhiên, không có mối tương quan rõ ràng giữa nồng độ Cd và các yếu tố thời tiết như nhiệt độ, độ ẩm và lượng mưa. Mối quan hệ giữa nồng độ Cd với các yếu tố trong đất như độ pH, độ ẩm và thành phần dinh dưỡng cho thấy, có ít tác động đáng kể đến sự tích tụ Cd trong đất. Phân tích sai số trong mô hình hồi quy tuyến tính nhấn mạnh sự cần thiết phải nâng cao độ chính xác và tin cậy của kết quả, đồng thời cho thấy, việc tối ưu hóa khả năng dự đoán nồng độ Cd không cần phụ thuộc vào các yếu tố thời tiết

TÀI LIỆU THAM KHẢO

1. Ren, J.H.; Sun, H.J.; Wang, S.F.; Luo, J.; Ma, L.Q. Interactive effects of mercury and arsenic on their uptake, speciation and toxicity in rice seedling. *Chemosphere* 2014, 117, 737-744.
2. Trung tâm quan trắc kỹ thuật - tài nguyên và môi trường Bình Dương. Báo cáo quan trắc đất 2022.
3. Wang, Z.Z.; Wang, H.B.; Wang, H.J.; Li, Q.C.; Li, Y. Effect of soil washing on heavy metal removal and soil quality: A two-sided coin. *Ecotoxicol Environ Saf.* 2020, 203, 110981.
4. Sutrisno, U., Wulandari, Y., Arifin, S., Mayeni Manurung, M., & Faisal, M. (2024). Trends, Contributions and Prospects: Bibliometric Analysis of ANOVA Research in 2022-2023 conditions of the Creative Commons Attribution-ShareAlike 4.0 International License-(CC-BY-SA). In *Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)*.
5. Petersen, F., & Hubbart, J. A. Spatial and temporal characterization of escherichia coli, suspended particulate matter and land use practice relationships in a mixed-land use contemporary watershed. *Water (Switzerland)* 2020, 12(5).