NGHIÊN CỨU ỨNG DỤNG KẾT HỢP CÔNG NGHỆ VIỄN THÁM VÀ THUẬT TOÁN HỌC MÁY MULTIPLE LINEAR REGRESSION TRONG THÀNH LẬP BẢN ĐỒ PHÁT THẢI BỤI MỊN $PM_{2.5}$

VŨ NGỌC PHAN[(1)], PHẠM MINH HẢI[(2)]

[(1)]Đại học Tài Nguyên và Môi Trường Hà Nội

[(2)]Viện Khoa học Đo đạc và Bản đồ

Tóm tắt:

Ô nhiễm môi trường không khí gây ra rất nhiều hậu quả cho con người. Chúng là tác nhân gây nên cái chết cho hàng triệu người mỗi năm. Theo WHO, ô nhiễm môi trường không khí gây ra 7 triệu ca tử vong mỗi năm, trong đó Châu Á - Thái Bình Dương chiếm khoảng 4 triệu ca. Trong đó, ô nhiễm bụi mịn $PM_{2.5}$ chính là thủ phạm gây ra nhiều ca tử vong nhất. Mục tiêu bài báo này là phát triển giải pháp thành lập bản đồ phát thải bụi mịn $PM_{2.5}$ ứng dụng kết hợp công nghệ viễn thám và thuật toán học máy Multiple Linear Regression. $PM_{2.5}$ là những hạt bụi li ti có trong không khí kích thước đường kính nhỏ hơn hoặc bằng 2.5 µm. Loại bụi này hình thành từ các chất như Carbon monoxide (CO), Sunphua điôxit (SO_2), Nitơ điôxit (NO_2) và các hợp chất kim loại khác, lơ lửng trong không khí. Việc tính toán $PM_{2.5}$ trong mối quan hệ tuyến tính giữa biến phụ thuộc $PM_{2.5}$ và các biến độc lập CO, SO_2, NO_2… (biến dự đoán) dựa trên thuật toán học máy Multiple Linear Regression có cơ sở khoa học và thực tiễn cao. Kết quả thực hiện của nghiên cứu này cung cấp giải pháp thành lập bản đồ phát thải bụi mịn $PM_{2.5}$ mang tính tự động hóa cao dựa vào số liệu viễn thám và các thông số quan trắc không khí mặt đất.

1. Mở đầu

Hiện nay, công tác quan trắc môi trường không khí tại các Tỉnh được thực hiện tại các trạm quan trắc môi trường không khí thuộc các Chi Cục bảo vệ môi trường hay Trung tâm Quan trắc Tài nguyên và Môi trường, Sở Tài Nguyên và Môi Trường các Tỉnh, Thành phố trực thuộc Trung ương. Trạm quan trắc cung cấp các chỉ số như: SO_2, NO_2, NO_x, CO, O_3, $PM_{2.5}$, PM_4, PM_{10}… và các chỉ số khí tượng như nhiệt độ, độ ẩm, tốc độ và hướng gió. Đây là cơ cơ sở đánh giá nhanh cũng như theo dõi diễn biến chất lượng môi trường. Tuy nhiên, do hạn chế về nguồn lực, số lượng các trạm quan trắc môi trường tự động tại các địa phương chưa nhiều, chưa theo kịp được với tình trạng ô nhiễm môi trường dưới tác động quá trình đô thị hóa nhanh và mạnh như hiện nay. Thành phố Hà Nội hiện có 11 trạm quan trắc môi trường tự động, thành phố Hồ Chí Minh có 9 trạm, Bắc Ninh có 16 trạm.v.v. Với số lượng hạn chế trạm quan trắc không khí tự động, việc nội suy chỉ số môi trường không khí cho khu vực rộng lớn cho toàn thành phố hay Tỉnh thường có độ chính xác chưa cao. Do đó, một giải pháp khoa học công nghệ góp phần nâng cao độ chính xác của các bản đồ chất lượng môi trường không khí có tính thời sự và cấp thiết cao.

Ngày nhận bài: 1/8/2023, ngày chuyển phản biện: 5/8/2023, ngày chấp nhận phản biện: 9/8/2023, ngày chấp nhận đăng: 28/8/2023

Trong phạm vi bài báo này, nhóm tác giả giới thiệu giải pháp thành lập bản đồ phát thải bụi mịn $PM_{2.5}$ ứng dụng kết hợp công nghệ viễn thám và thuật toán học máy Multiple Linear Regression. $PM_{2.5}$ là những hạt bụi li ti có trong không khí kích thước đường kính nhỏ hơn hoặc bằng 2.5 µm. Việc tính toán $PM_{2.5}$ trong mối quan hệ tuyến tính giữa biến phụ thuộc $PM_{2.5}$ và biến độc lập CO, SO_2, NO_2 … (biến dự đoán) dựa trên thuật toán học máy Multiple Linear Regression có cơ sở khoa học và thực tiễn cao.

2. Khu vực thực hiện và dữ liệu đầu vào

2.1. Khu vực thực hiện

Bắc Ninh là tỉnh có diện tích nhỏ nhất cả nước, nằm cách trung tâm Thành phố Hà Nội 30 km về phía đông bắc. Với vị trí nằm trong Vùng thủ đô Hà Nội, vùng kinh tế trọng điểm Bắc Bộ, thuộc vùng Đồng bằng sông Hồng, Bắc Ninh có nhiều điều kiện thuận lợi trong phát triển kinh tế. Tỉnh có diện tích là 822,71 km² và số dân là 1.488.250 người. Tỉnh có địa hình không hoàn toàn là đồng bằng mà xen kẽ là các đồi thấp có hướng dốc chủ yếu từ Bắc xuống Nam và từ Tây sang Đông, được thể hiện qua các dòng chảy bề mặt đổ về sông Đuống và sông Thái Bình. Vùng đồng bằng thường có độ cao phổ biến từ 3-7 m, địa hình trung du (thị xã Quế Võ và huyện Tiên Du) có một số dải đồi thấp độ cao không quá 200 m.

Hình 1: Bản đồ hành chính Tỉnh Bắc Ninh

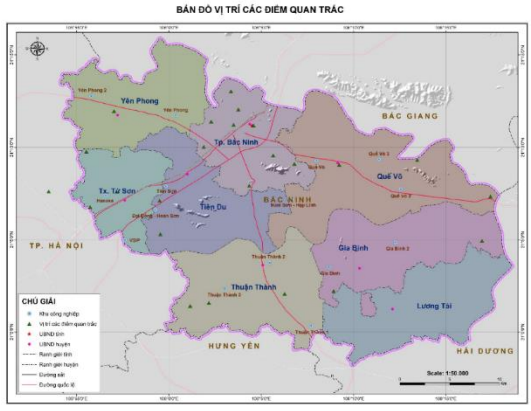
2.2. Dữ liệu đầu vào

a. Ảnh vệ tinh Landsat OLI8

Nhóm nghiên cứu thử nghiệm dữ liệu ảnh vệ tinh Landsat OLI8 chụp ngày 01/11/2022 tại Tỉnh Bắc Ninh có độ phân giải mặt đất là 30m. Dữ liệu được tải miễn phí từ website: https://earthexplorer.usgs.gov/. Ảnh vệ tinh trong đề tài có độ phủ mây nhỏ hơn 10%, và được hiệu chỉnh bức xạ và khí quyển bằng công cụ ATCOR (Atmospheric correction) trong phần mềm PCI Geomatic 2018. Ảnh được sử dụng hệ tọa độ WGS84 và hệ quy chiếu UTM múi 48.

b. Dữ liệu quan trắc mặt đất

Nhóm nghiên cứu đã sử dụng dữ liệu quan trắc mặt đất tại 7 trạm quan trắc tự chế tạo và 16 trạm quan trắc môi trường tự động của Trung tâm Quan trắc Tài nguyên và Môi trường, Sở Tài Nguyên và Môi Trường Tỉnh Băc NinhMỗi bản ghi trong tập dữ liệu chứa các cột: Nox, SO_2, O_3, PM_{10}, $PM_{2.5}$, CO, áp suất, nhiệt độ, hướng gió.v.v. Thời gian lấy mẫu cách nhau trung bình khoảng 1 giờ đồng hồ. Tuy nhiên, tập dữ liệu tồn tại một số bản ghi có giá trị rỗng và bị nhiễu. Biểu đồ phân bố các giá trị của thuộc tính (các cột) được mô tả trong Hình 2.



Hình 2: Sơ đồ vị trí các điểm quan trắc

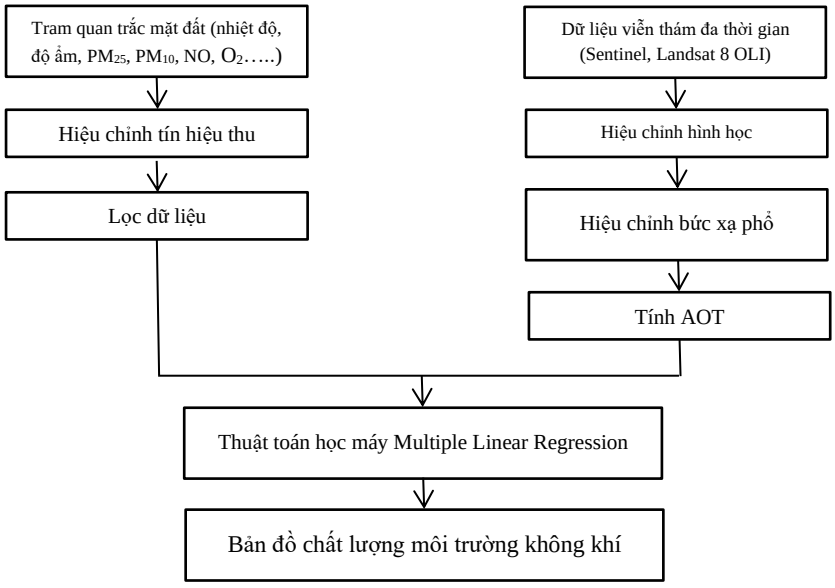
Bởi dữ liệu được thu thập được lấy mẫu cách nhau 1 giờ đồng hồ, tuy nhiên do thời

gian chụp ảnh Landsat vào 10h sáng nên nhóm nghiên cứu lấy 1 giá trị/ngày tính trung bình các chỉ số trên các bản ghi từ 6 đến 12 giờ. Kết quả thu được 11.000 bản ghi về các chỉ số không khí. Tiếp theo, chúng tôi thực hiện tiền xử lý, trích rút và chuẩn hóa dữ liệu này. Để thực hiện quá trình huấn luyện và đánh giá, các bản ghi được lấy ngẫu nhiên và chia thành 2 tập: tập huấn luyện (training set) chiếm 75% dữ liệu ban đầu và 25% dữ liệu còn lại là tập kiểm tra (test set).

3. Cơ sở khoa học kết hợp công nghệ viễn thám và thuật toán học máy Multiple Linear Regression trong thành lập bản đồ phát thải bụi mịn $PM_{2.5}$

3.1. Quy trình công nghệ kết hợp công nghệ viễn thám và thuật toán học máy Multiple Linear Regression trong thành lập bản đồ phát thải bụi mịn $PM_{2.5}$

Sau khi thu thập dữ liệu về các chỉ số quan trắc môi trường, nhóm nghiên cứu tiến hành xây dựng Quy trình công nghệ kết hợp công nghệ viễn thám và thuật toán học máy Multiple Linear Regression trong thành lập bản đồ phát thải bụi mịn $PM_{2.5}$ (Hình 2).



Hình 3: Quy trình công nghệ kết hợp công nghệ viễn thám và thuật toán học máy Multiple Linear Regression trong thành lập bản đồ phát thải bụi mịn $PM_{2.5}$

3.2. Tiền xử lý ảnh Landsat 8OLI

Việc tiền xử lý ảnh được tiến hành bằng cách chuyển các giá trị số (DN - Digital Nember) sang giá trị bước xạ phổ hoặc phản xạ phổ. Có nhiều mức hiệu chỉnh bức xạ. Đầu tiên chuyển đổi DN thành giá trị bức xạ tại đầu thu, thứ hai là chuyển đổi bức xạ phổ tại đầu thu về bức xạ phổ ở bề mặt trái đất, cuối cùng tiến hành hiệu chỉnh khí quyển ảnh để loại bỏ ảnh hưởng của điều kiện khí quyển đến chất lượng ảnh.

Chuyển đổi DN sang giá trị bức xạ phổ tại đỉnh khí quyển (TOA):

Dữ liệu ảnh Landsat 8 OLI được chuyển đổi sang dữ liệu bức xạ phổ đỉnh khí quyển [4] sử dụng công thức sau:

$$L_\lambda = M_L * Q_{cal} + A_L \quad (1)$$

Trong đó:

L_λ - Bức xạ phổ đỉnh khí quyển (Watts/(m^2 * srad * μm)

M_L - Hệ số thay đổi tỷ lệ bức xạ của kênh ảnh theo tính chất đa bội, được lấy trong tệp dữ liệu metadata (RADIANCE_MULT_BAND_x, trong đó x là kênh ảnh)

A_L - Hệ số thay đổi tỷ lệ bức xạ của kênh ảnh theo tính chất cộng dồn, được lấy trong tệp dữ liệu metadata (RADIANCE_ADD_BAND_x, trong đó x là kênh ảnh)

Qcal - Lượng tử hóa và hiệu chuẩn tiêu chuẩn giá trị số của kênh ảnh (DN)

Dữ liệu các kênh ảnh Landsat 8 OLI chuyển đổi thành phản xạ tại đỉnh khí quyển TOA bằng cách sử dụng hệ số phản xạ hồi quy được cung cấp trong tệp dữ liệu matadata (tệp tin MTL). Phương trình sau đây được sử dụng để chuyển đổi các giá trị DN sang phản xạ TOA đối với dữ liệu Landsat 8 OLI [4] như sau:

$$\rho\lambda' = M_\rho * Q_{cal} + A_\rho \quad (2)$$

Trong đó:

ρλ' - TOA Phản xạ tại đỉnh khí quyển, chưa hiệu chỉnh góc tới

Mρ - Hệ số thay đổi tỷ lệ phản xạ của kênh ảnh

Aρ - Hệ số thay đổi tỷ lệ phản xạ của kênh ảnh theo tính chất cộng dồn, được lấy trong tệp dữ liệu metadata

Qcal - Lượng tử hóa và hiệu chuẩn tiêu chuẩn giá trị số của kênh ảnh (DN)

TOA phản xạ đỉnh khí quyển khi hiệu chỉnh góc tới mặt trời:

$$\rho\lambda = \frac{\rho\lambda'}{\cos(\theta_{SZ})} = \frac{\rho\lambda'}{\sin(\theta_{SE})} \quad (3)$$

Trong đó:

ρλ - TOA Phản xạ tại đỉnh khí quyển

θ_{SE} - Góc tới mặt trời (SUN_ELEVATION).

θ_{SZ} - Góc thiên đỉnh mặt trời; $\theta_{SZ} = 90° - \theta_{SE}$

3.3. Lọc dữ liệu bằng phương pháp Pearson Correlation

Nhóm nghiên cứu tiến hành loại bỏ các bản ghi nhiễu, bị khuyết, mang giá trị nằm ngoài miền cho phép ví dụ như: không có giá trị hay -9999. Qua khảo sát, các yếu tố về khí tượng như: lượng mưa, độ ẩm, nhiệt độ được nhóm nghiên cứu giữ lại bởi các chỉ số này phản ánh về điều kiện thời tiết và môi trường. Chúng cũng là nhân tố quan trọng trong dự báo ô nhiễm bụi $PM_{2.5}$. Tiếp theo nhóm nghiên cứu đã hệ số tương quan theo phương pháp Pearson và biểu đồ heatmap để loại bỏ những chỉ số không cần thiết trong việc tính toán hàm lượng bụi $PM_{2.5}$. Có một vài kiểu tính hệ số tương quan như Pearson, Kendall, hay Spearman nhưng phương pháp phổ biến nhất là Pearson correlation (r). Phương pháp này đo lường độ mạnh và hướng của mối quan hệ tuyến tính giữa hai biến, không thể áp dụng cho hai biến không có mối quan hệ tuyến tính và cũng không thể phân biệt được biến độc lập và biến phụ thuộc. Hệ số tương quan theo phương pháp Pearson Correlation được biểu thị theo công thức dưới đây:

$$\rho_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \quad (8)$$

Trong đó:

ρ_{xy}: Hệ số tương quan Pearson

Cov(x,y): Hiệp phương sai của biến x và y

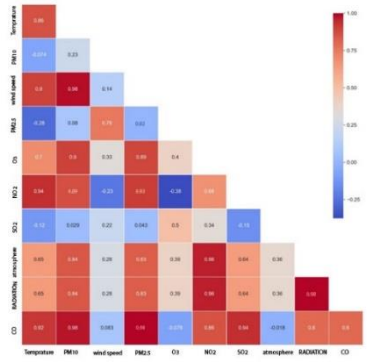
σ_x: Độ lệch chuẩn của biến x

σ_y: Độ lệch chuẩn của biến y

Heatmap là biểu đồ sử dụng cường độ màu sắc để thể hiện độ lớn của giá trị. Khi đó các giá trị lớn sẽ được làm nổi bật bằng các

vùng màu có cường độ ánh sáng mạnh và các giá trị nhỏ hơn sẽ được thể hiện bằng các mảnh màu nhạt hơn. Những vùng có trọng số lớn thường được thể hiện là màu đỏ, những vùng có trọng số nhỏ hơn sẽ có sắc độ nhiệt giảm dần từ cam - vàng - xanh lá đến xanh da trời. Qua đó, chúng ta sẽ phân biệt được những yếu tố có trọng số lớn trong mối tương quan với các yếu tố khác đang sử dụng để thống kê. Từ đó quyết định được việc sử dụng yếu tố nào trong toàn bộ các yếu tố đang sử dụng để so sánh. Nhóm thực hiện đề tài sử dụng hàm Correlation trong thư viện của ngôn ngữ lập trình python trong tính toán biểu diễn hệ số tương quan các chỉ số không khí.

Nhìn vào hình 4, chúng ta có thể thấy sự tương quan của $PM_{2.5}$ với CO và NO_2 cao (0.98 và 0.93), trong khi sự tương quan với SO_2, áp suất khí quyển, ánh sáng thấp hơn (0.043, 0.83, 0.83). Do đó, bốn chỉ số SO_2, áp suất khí quyển, ánh sáng, gió được loại bỏ trong mô hình tính hàm lượng bụi $PM_{2.5}$ sau này.



Hình 4: Mối tương quan của biến độc lập và phụ thuộc trên biểu đồ Heatmap

3.4. Độ dày quang học khí quyển AOT

Độ dày quang học khí quyển (AOT) là một chỉ số của tải lượng sol khí trong cột khí quyển theo chiều thẳng đứng từ bề mặt đến đỉnh của tầng khí quyển. Độ dày quang học của khí quyển (AOT) là đại lượng đặc trưng cho mức độ suy giảm bức xạ do hấp thụ và tán xạ bức xạ mặt trời của sol khí. Giá trị AOT càng lớn thì khí quyển càng vẩn đục hay nồng độ sol khí nhiều. Dựa trên sự suy giảm năng lượng tới đầu thu vệ tinh do bị hấp thụ, tán xạ của các phân tử khí ô nhiễm và các hạt bụi từ đấy tính toán hàm lượng bụi trong không khí. Sau khi hiệu chỉnh khí quyển, ta tính được phản xạ ở đỉnh của khí quyển (TOA) và phản xạ mặt đất từ đó ta tính được phản xạ khí quyển. Từ đó, tính độ dày sol khí (AOT) như sau được đưa ra bởi [2]:

$$AOT(\lambda) = a_0 R(\lambda) \qquad (9)$$

Trong đó:

R(λ) - Hàm phản xạ khí quyển tương ứng với bức sóng (λ)

Phương trình trên được viết lại cho các kênh ảnh như sau:

$$AOT(\lambda) = a_0 R_{\lambda 1} + a_j R_{\lambda 2} + a_2 R_{\lambda 3} + ... \qquad (10)$$

Trong đó Rλi là phản xạ khí quyển (i = 1, 2 và 3 tương ứng với bước sóng vệ tinh), và a_j là hệ số thuật toán (j = 0, 1 và 2) được xác định bằng thực nghiệm.

3.5. Thuật toán học máy Multiple Linear Regression

Multiple Regression là một thuật toán hồi quy tuyến tính đa biến trong Machine Learning. Nó dùng để dự đoán một giá trị của biến phụ thuộc y dựa vào 2 hoặc nhiều biến độc lập x_1, x_2, x_3…x_n. Dạng tổng quát của thuật toán Multiple Regression có phương trình như sau:

$$y_i = w_0 + w_1 x_{i1} + w_1 x_{i2} + w_1 x_{i3} + ... + w_1 x_{ip} = w_T x_i \qquad (11)$$

Trong đó:

yi: là biến phụ thuộc

x_{i1}, x_{i2}, x_{i3}…x_{ip}: là các biến độc lập

w_0: là hằng số

w_1, w_2, w_3,w_n: là các hệ số quan hệ

Gọi x_i là một véc tơ đại diện cho quan sát thứ i, các giá trị cụ thể tương ứng là (x_{i1}, x_{i2}, x_{i3}…x_{ip}). Ma trận X có kích thước n x p có mỗi hàng là một quan sát và mỗi cột là một biến. Giá trị xip là quan sát thứ i của biến thứ p. Gọi ma trận mở rộng của ma trận X là ma trận có thêm véc tơ cột 1 được thêm vào ở cột đầu tiên.Khi đó các tập dữ liệu được biểu diễn như sau:

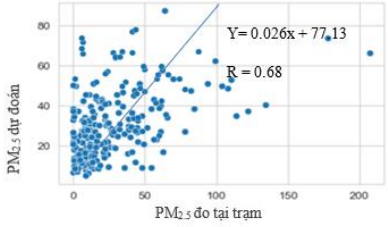
$$y = f(X) = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix} = \overline{X}w \quad (12)$$

Nghiệm của phương trình hồi quy là:

$$w = \left(\overline{X}^T \overline{X}\right)^{-1} \overline{X}^T y = (A^{-1}b) \quad (13)$$

$$A = \overline{X}^T \overline{X} \text{ và } \overline{X}^T y = b \quad (14)$$

Biểu đồ tương quan tuyến tính giữa giá trị $PM_{2.5}$ tính và giá trị $PM_{2.5}$ đo tại trạm như sau:



Hình 5: Biểu đồ tương quan tuyến tính giữa giá trị $PM_{2.5}$ dự đoán và giá trị $PM_{2.5}$ đo tại trạm

Kết quả đánh giá tương quan cho kết quả như sau:

Method:	Least Squares	F-statistic:	77.13
Date:	Tue, 23 May 2023	Prob (F-statistic):	8.60e-44
Time:	16:05:25	Log-Likelihood:	-3779.5
No. Observations:	798	AIC:	7567.
Df Residuals:	794	BIC:	7586.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-7.0125	4.039	-1.736	0.083	-14.941	0.916
x1	0.0510	0.018	2.844	0.005	0.016	0.086
x2	0.5848	0.068	8.551	0.000	0.451	0.719
x3	0.0244	0.003	7.809	0.000	0.018	0.031

Omnibus:	371.199	Durbin-Watson:	1.565
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3137.814
Skew:	1.909	Prob(JB):	0.00
Kurtosis:	11.933	Cond. No.	2.78e+03

Công thức tính $PM_{2.5}$ được mô tả dưới hình sau:

$$PM_{2.5} = -7.0125 + 0.051 * DN + 0.58 * NO_2 + + 0.024 * CO \quad (15)$$

3.6. Thuật toán nội suy CO và NO_2

Nghiên cứu đã sử dụng nội suy Inverse Distance Weighting (IDW). Đây là một trong những kỹ thuật phổ biến nhất để nội suy các điểm phân tán. Phương pháp IDW xác định giá trị của các điểm chưa biết bằng cách tính trung bình trọng số khoảng cách các giá trị của các điểm đã biết giá trị trong vùng lân cận của mỗi pixel. Những điểm càng cách xa điểm cần tính giá trị càng ít ảnh hưởng đến giá trị tính toán, các điểm càng gần thì trọng số càng lớn. Phương pháp nội suy định lượng khoảng cách ngược cho rằng mỗi điểm đầu vào có những ảnh hưởng cục bộ làm rút ngắn khoảng cách. Phương pháp này tác dụng vào những điểm ở gần điểm đang xét hơn so với những điểm ở xa. Số lượng các điểm chi tiết, hoặc tất cả những điểm nằm trong vùng bán kính xác định có thể được sử dụng để xác định giá trị đầu ra cho mỗi vị trí. Trọng số của mỗi điểm được tính theo công thức sau:

$$Z_0 = \frac{\sum_{i=1}^{N} Z_1 \times d_1^{-n}}{\sum_{i=1}^{N} d_i^{-n}} \quad (16)$$

Trong đó:

- Z0: giá trị ước tính của biến z tại điểm i.

- Zi: giá trị mẫu tại điểm i.

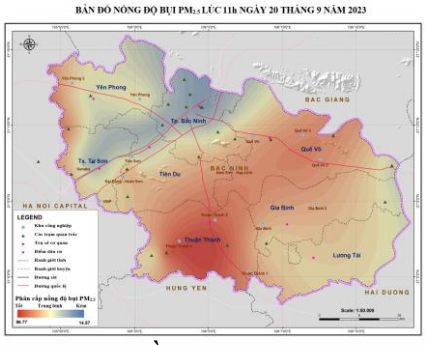
- D1: khoảng cách điểm mẫu để ước tính điểm.

- N: hệ số xác định trọng lượng dựa trên một khoảng cách

4. Kết quả thực nghiệm thành lập bản đồ phát thải bụi mịn $PM_{2.5}$ kết hợp công nghệ viễn thám và thuật toán học máy Multiple Linear Regression

4.1. Kết quả tính AOT

Trên cơ sở phương trình trên, nhóm nghiên cứu tiến hành khảo sát phân tích tương quan và hồi quy các mô hình tính bụi $PM_{2.5}$ với ảnh Landsat 8 OLI. Kết quả tính toán AOT sau đó được sử dụng trong thành lập bản đồ phát thải bụi mịn $PM_{2.5}$ bằng thuật toán (15). Kết quả thực nghiệm thành lập bản đồ phát thải bụi mịn $PM_{2.5}$ kết hợp công nghệ viễn thám và thuật toán học máy Multiple Linear Regression ở hình dưới.



Hình 9: Bản đồ phát thải bụi mịn $PM_{2.5}$

Do nghiên cứu sử dụng ảnh Landsat OLI 8 với giờ chụp ảnh tại khu vực nghiên cứu từ 10-11h sáng do đó, các thông số quan trắc mặt đất cũng được lấy trung bình tại hai khung giờ này đảm bảo sự thống nhất về mặt thời gian. Dựa vào bản đồ phát thải khí $PM_{2.5}$ trên cho thấy, phần lớn khu vực trên địa bàn tỉnh Bắc Ninh nằm trong mức 0-100, đây là mức chất lượng không khí từ trung bình đến kém. Chúng ta thấy nồng độ $PM_{2.5}$ cao tập trung tại các khu vực như Huyện Yên Phong, Huyện Quế Võ, Thành phố Bắc Ninh, phía bắc của Huyện Tiên Du và một phần của Huyện Thuận Thành, nồng độ bụi giảm dần qua những huyện ngoại thành khác. Sự khác biệt này có thể được đánh giá qua sự phát triển về giao thông, phân bố làng nghề và công nghiệp chênh lệch giữa các huyện, nhóm các huyện có kết quả phát thải bụi mịn $PM_{2.5}$ thấp hơn là những huyện vẫn còn những diện tích lớn vùng phát triển nông nghiệp, những huyện và thành phố còn lại thuộc nhóm đi đầu về phát triển công nghiệp cả trong và ngoài tỉnh Bắc Ninh. Một phần diện tích của Thành phố Bắc Ninh nơi tiếp xúc với Huyện Quế Võ thể hiện rất kém (=100). Đây là mức chất lượng không khí xấu (nhóm nhạy cảm tránh ra ngoài, những người khác nên hạn chế thời gian ở ngoài). Kết quả nghiên cứu này cũng tiếp tục đưa ra cảnh báo về ô nhiễm bụi siêu mịn trên địa bàn tỉnh Bắc Ninh bởi bụi có kích thước 2.5 µm rất dễ dàng xâm nhập vào cơ thể qua đường hô hấp và qua da, gây ảnh hưởng tiêu cực đến sức khỏe, đặc biệt là nguy cơ mắc các bệnh về đường hô hấp, hệ thần kinh và não bộ, có thể gây ung thư và biến đổi gien.

5. Kết luận

Nghiên cứu đã thử nghiệm ứng dụng kết hợp công nghệ viễn thám, thuật toán học máy Multiple Linear Regression, và số liệu quan trắc mặt đất trong thành lập bản đồ phát thải bụi mịn $PM_{2.5}$ với khu vực thử nghiệm tại Bắc Ninh. Multiple Regression là một thuật toán hồi quy tuyến tính đa biến trong Machine Learning. Nó dùng để dự đoán một giá trị của biến phụ thuộc y dựa vào 2 hoặc nhiều biến độc lập. Việc tính toán $PM_{2.5}$ trong mối quan hệ tuyến tính giữa biến phụ thuộc $PM_{2.5}$ và biến độc lập CO, SO_2, NO_2… (biến dự đoán) dựa trên thuật toán học máy Multiple Linear Regression có cơ sở khoa học và thực tiễn cao. Nghiên cứu đã phát triển được thuật toán tính toán $PM_{2.5}$ (15) dựa trên thuật toán học máy Multiple Linear Regression. Kết quả thực hiện đã thành lập được bản đồ bụi mịn $PM_{2.5}$ tại khu vực thử nghiệm. Qua đó, chúng ta thấy nồng độ $PM_{2.5}$ cao tập trung tại các khu vực như Huyện Yên Phong, Huyện Quế Võ, Thành phố Bắc Ninh, phía bắc của Huyện Tiên Du và một phần của Huyện Thuận Thành, nồng độ bụi giảm dần qua những huyện ngoại thành khác. Phương hướng nghiên cứu tiếp theo, nhóm nghiên cứu sẽ tiến hành thực nghiệm thêm các thuật toán AI

khác để có cơ sở kiểm nghiệm, lựa chọn thuật toán AI phù hợp trong tính toán thành lập bản đồ phát thải khí $PM_{2.5}$ sử dụng dữ liệu viễn thám và các trạm quan trắc mặt đất trong tương lai.❍

Lời cảm ơn:

Nhóm thực hiện nghiên cứu xin chân thành cảm ơn Bộ Tài Nguyên và Môi Trường đã tài trợ thực hiện đề tài nghiên cứu khoa học cấp Bộ: "Nghiên cứu ứng dụng trí tuệ nhân tạo cho dự báo, cảnh bảo chất lượng môi trường không khí theo số liệu viễn thám, các trạm quan trắc môi trường mặt đất" mã số: TNMT.2022.04.06.

Tài liệu tham khảo

[1]. Lim, H., et al., Remote sensing of PM10 from LANDSAT TM imagery. Acrs 2004, 2004: p. 739-744. 8.

[2]. Nadzri, O., Z.M.J. Mohd, and H.S. Lim, Estimating Particulate Matter Concentration over Arid Region Using Satellite Remote Sensing: A Case Study in Makkah, Saudi Arabia. . Modern applied Science 4., 2010: p. 131-142. 9.

[3]. Sam Appadurai.A and J.Colins JohnnyM.E, Satellite based estimation of pm10 from AOT of landsat 7ETM+ over Chennai city. International Journal of Advances in Engineering Research, 2016. Vol. No. 11.

[4]. Using the USGS Landsat 8 Product, https://landsat.usgs.gov/Landsat8_Using_Product.php 17.

[5]. Moran, M.S., et al., Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. Remote Sensing of Environment, 1992. 41(2-3): p. 169-184.

[6]. Chavez, P.S., Image-based atmospheric corrections-revisited and improved. Photogrammetric engineering and remote sensing, 1996. 62(9): p. 1025-1035.

[7]. Sobrino, J.A., J.C. Jiménez-Muñoz, and L. Paolini, Land surface temperature retrieval from LANDSAT TM 5. Remote Sensing of Environment, 2004. 90(4): p. 434-440.❍

Summary

Application of remote sensing and Multiple Linear Regression AI algorithm in mapping $PM_{2,5}$

Vu Ngoc Phan, Hanoi University of Natural Resources and Environment

Pham Minh Hai, The Viet Nam Institute of Surveying and Mapping (VISAM)

Air pollution causes many problems for humans around the world. According to the WHO, air pollution causes 7 million deaths annually, of which Asia-Pacific accounts for about 4 million. In particular, fine dust pollution $PM_{2.5}$ is the culprit causing the most deaths. The research objective is to develop a solution to establish a $PM_{2.5}$ map using remote sensing and a Multiple Linear Regression AI algorithm. $PM_{2.5}$ are tiny dust particles in the air with a diameter of less than or equal to 2.5 µm. This type of dust is formed from substances such as Carbon monoxide (CO), Sulfur dioxide (SO_2), Nitrogen dioxide (NO_2) and other metal compounds suspended in the air. The calculation of $PM_{2.5}$ in the linear relationship between the dependent variable $PM_{2.5}$ and the independent variables CO, SO_2, NO_2... (predictor variable) is based on the Multiple Linear Regression machine learning algorithm with a highly scientific and practical mean. The results of this study provide a solution to create an automated $PM_{2.5}$ fine dust map based on remote sensing data and ground air monitoring parameters.❍