**Mục tiêu:**

Xử lý dữ liệu dòng chảy lịch sử, tính toán các thống kê và xu hướng, và đánh giá ảnh hưởng của hiện tượng El Niño đến dòng chảy thông qua mô hình hồi quy và trực quan hóa.

**Các tệp tin:**

- TH2\_Input\_22166022.xlsx – **Dữ liệu dòng chảy thô của một trạm quan trắc tại Thái Lan, mỗi sheet tương ứng với một năm.**

- [TH2\_Output\_22166022.xlsx](https://github.com/Hanna137/Data-mining/blob/main/TH2_Output_22166022.xlsx" \o "TH2_Output_22166022.xlsx): chuyển input từ bảng ngang sang dạng bảng dọc , lưu lượng dòng chảy theo ngày.

**-** [avg.xlsx](https://github.com/Hanna137/Data-mining/blob/main/avg.xlsx" \o "avg.xlsx): kết quả tính trung bình dòng chảy tháng của các năm

- [stats.xlsx](https://github.com/Hanna137/Data-mining/blob/main/stats.xlsx" \o "stats.xlsx): bảng thống kê mô tả cơ bản cho các tháng

- nino34.long.anom.data.xlsx – **Dữ liệu chỉ số NINO 3.4 (El Niño).**

- code.R: tất cả code của bài.

- result.docx: file kết quả chứa biểu đồ.

**Kết quả phân tích ảnh hưởng của thời gian và El NiNo/La NiNa đến dòng chảy trung bình:**

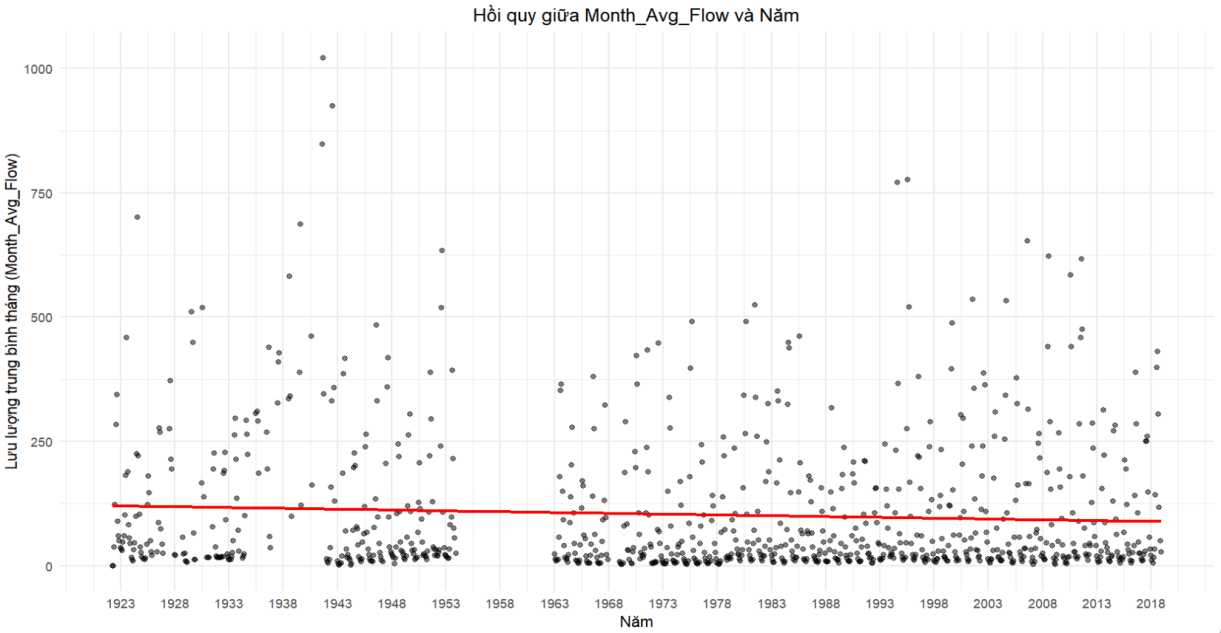
**(Hình 1)** Dựa trên biểu đồ hồi quy giữa Month\_Avg\_Flow và Năm, ta nhận thấy:

- Xu hướng tổng thể của dòng chảy trung bình tháng giảm nhẹ từ 1922-2018, thể hiện qua đường hồi quy màu đỏ có hệ số dốc âm.

- Sự phân tán dữ liệu rất lớn, đặc biệt là những năm có lưu lượng cao đột biến, cho thấy biến thiên mạnh theo năm.

- Khoảng trắng ở giữa biểu đồ là do thiếu dữ liệu của các năm từ 1954-1962, có nhiều giá trị ngoại vi đặc biệt là ở các năm 1940, 1980, 2000.

- Tuy nhiên, đường xu hướng khá phẳng, cho thấy tốc độ giảm không lớn.

****

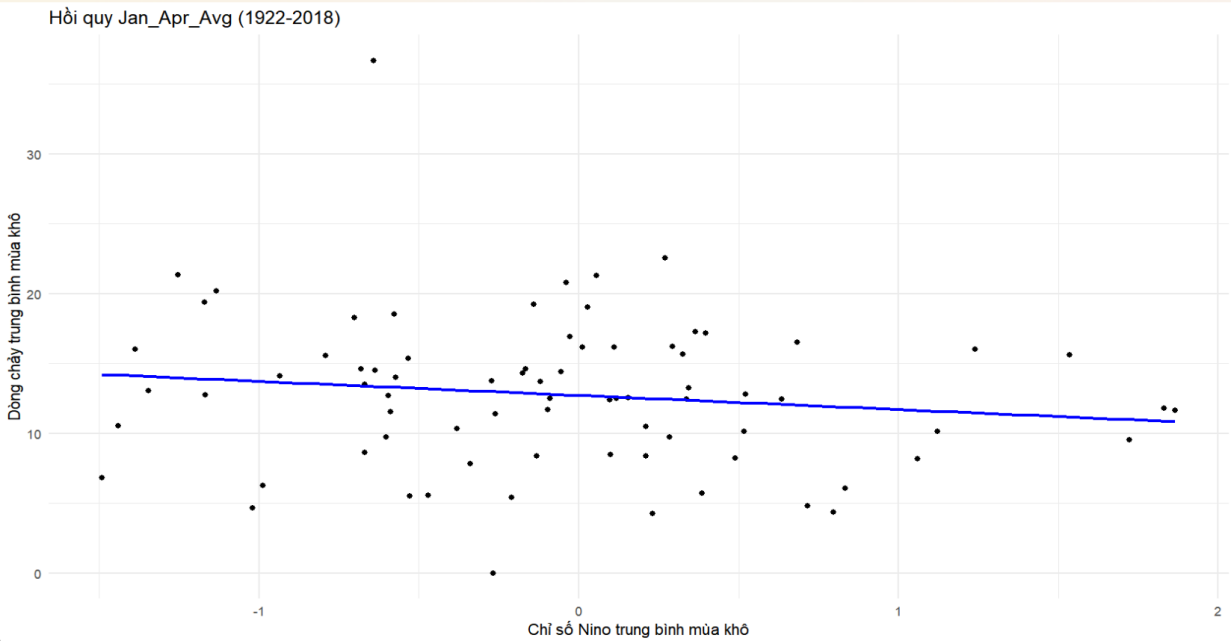
***Hình 1. Hồi quy giữa trung bình các tháng và năm***

**(Hình 2)** Dòng chảy mùa khô (Jan–Apr)

- Xu hướng giảm nhẹ của dòng chảy theo chỉ số Nino.

- Khi chỉ số Nino tăng (El Niño), dòng chảy mùa khô có xu hướng giảm.

- Điều này phù hợp với lý thuyết: El Niño thường gây ít mưa hơn ở một số khu vực Đông Nam Á trong mùa khô ( cụ thể ở đây là Thái Lan).

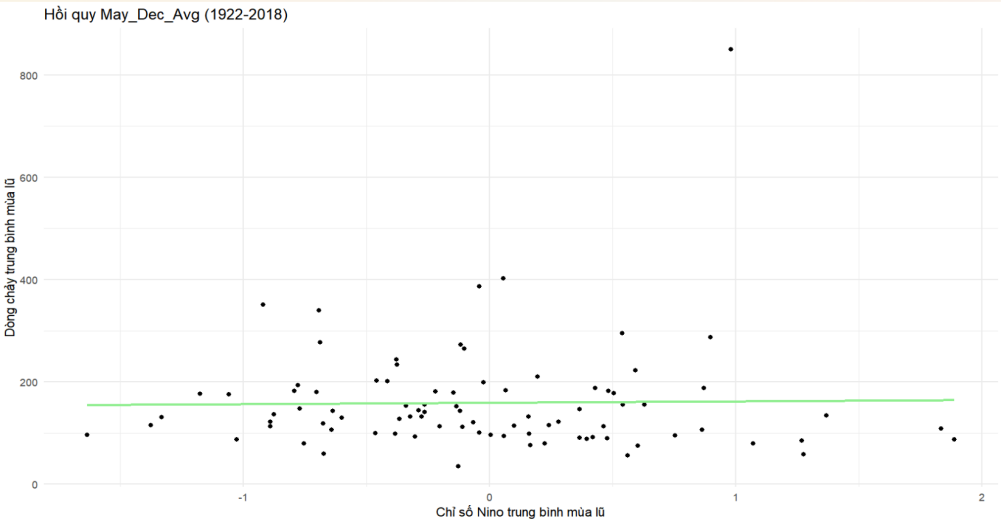


***Hình 2. Hồi quy từ tháng 1 đến tháng 4 năm 1922 -2018***

**(Hình 3)** Dòng chảy mùa lũ (May–Dec)

- Đường hồi quy gần như không có xu hướng rõ ràng. Đường hồi quy khá ngang, cho thấy không có mối quan hệ mạnh giữa Nino và dòng chảy mùa lũ ở khu vực này.

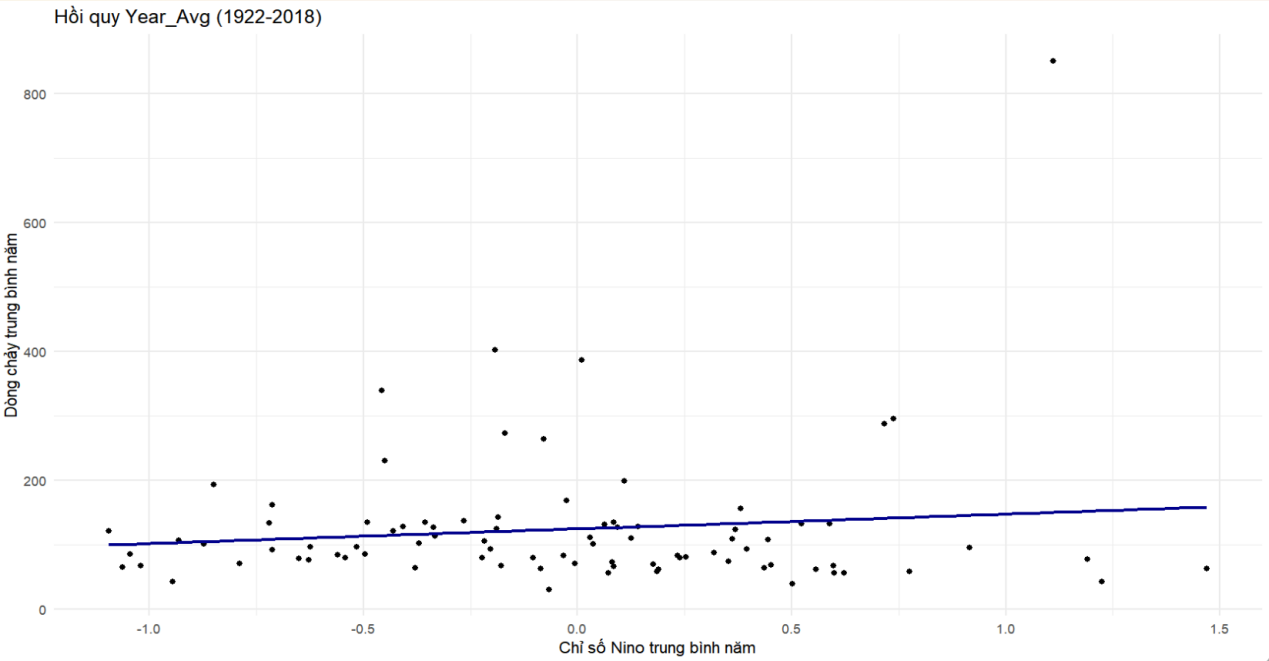
- Dòng chảy dao động mạnh từ dưới 100 đến hơn 800, không theo trật tự rõ ràng nào, có năm dòng chảy rất cao nhưng không trùng với giá trị Nino.



***Hình 3. Hồi quy từ tháng 5 đến tháng 12 năm 1922-2018***

**(Hình 4)** Dòng chảy trung bình năm

- Có xu hướng tăng nhẹ theo chỉ số Nino, dữ liệu phân tán khá mạnh, có nhiều điểm dữ liệu lệch xa, đặc biết là những năm có dòng chảy trên 400m3/s.

- Đây là dấu hiệu cho thấy khu vực này có thể bị ảnh hưởng bởi các hiện tượng khí hậu toàn cầu như El Niño/La Niña, dù ảnh hưởng yếu.

***Hình 4. Hồi quy trung bình các năm 1922-2018***