

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**  
**TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT**

**BÁO CÁO SINH HOẠT HỌC THUẬT KỲ II NĂM HỌC 2022-2023**

**TÊN BÁO CÁO:** Địa thống kê và bản đồ trường Địa vật lý

**BÁO CÁO VIÊN:** PGS.TS Phan Thiên Hương

**Hà Nội, 5/2023**

## MỞ ĐẦU

Đại đa số các nhà khoa học trái đất thường không đi sâu vào toán học, thí dụ như địa chất truyền thống thường mang tính định tính. Tuy nhiên với việc thu thập số liệu ngày càng nhiều, những tính chất ban đầu được cho là ngẫu nhiên, các số liệu dựa trên các sơ đồ phân loại và mô tả liên quan đến các hiện tượng vật lý dần dần lại thể hiện tính quy luật của nó dưới con mắt của các nhà toán học thống kê. Vì vậy địa thống kê ngày càng phát triển và dần không thể thiếu được trong các kết quả nghiên cứu của các nhà khoa học nghiên cứu đất. Trong cuốn sách của mình, *Mô hình ngẫu nhiên và địa thống kê* (1994), Yarus thảo luận về những gì ông tin là vấn đề cốt lõi trong mô hình địa chất của các bể (geological reservoir).

Nhiều nhà địa chất học dùng quen các khái niệm phần lớn theo *định tính*. Thuật ngữ định tính này có giá trị trong việc cho phép các nhà địa chất mô tả các bề mặt cấu trúc như các cấu trúc lồi, lõm. Môi trường lắng đọng được mô tả dưới dạng như đồng bằng châu thổ, rạn san hô hoặc quạt biển sâu, cùng với các bề mặt chùng lún hoặc trượt xuống. Cuối cùng, thạch học là đá sa thạch, đá vôi và đá phiến sét. Mặc dù mang tính mô tả, những thuật ngữ này cho phép ý nghĩa của chúng có phạm vi rộng lớn và tùy thuộc vào cách giải thích của từng cá nhân. Chất lượng diễn giải phần lớn phụ thuộc vào kinh nghiệm của người diễn giải, mặc dù chúng ta mong chờ các nhà địa chất khác tạo ra các kết quả tương tự nhưng hơi khác một chút, cuối cùng thì kết luận đều có thể là các mô hình bể chứa tiềm năng.

Tuy nhiên như đã nói ở trên với một số lượng thông tin lớn thì dần dần các thông tin này sẽ thể hiện tính quy luật của chúng, như các nhà toán học thống kê nói thì chúng sẽ đwocj thể hiện bởi những con số, tính định tính giảm dần và tính định lượng sẽ tăng lên. Trong báo cáo này tôi không có tham vọng đi hết tất cả các thuật toán thống kê xác suất liên quan đến khoa học trái đất mà chỉ mong muốn trình bày một số khái niệm cơ bản và nhưng đặc trưng liên quan đến xây dựng bản đồ địa vật lý.

Ở đây đôi khi chúng ta tạm hiểu bể là ẩn định một không gian có giới hạn. Đặc biệt trong dầu khí sự quan tâm lớn nhất của họ chính là bể chứa dầu khí, các tính chất liên quan đến bể chính là sự thay đổi các tham số vật lý trong không gian giới hạn của bể.

### I. KHÁI NIỆM CƠ BẢN

*Địa thống kê* có thể được định nghĩa là một nghiên cứu về mối tương quan không gian giữa các biến. Nhánh toán học và thống kê ứng dụng đang phát triển nhanh chóng này cung cấp một tập hợp các công cụ nhằm tìm hiểu và lập mô hình *biến đổi không gian*.

Tính biến đổi theo không gian bao gồm các thang đo của tính không đồng nhất cũng như *tính định hướng* trong các tập dữ liệu. Dữ liệu bể chứa thể hiện khả năng kết nối không gian ở mức độ lớn hơn hoặc thấp hơn, bởi vì khi khoảng *cách* giữa hai điểm dữ liệu tăng

lên, độ tương tự giữa hai phép đo giảm đi. Và bởi vì các bể chứa có xu hướng thể hiện tính dị hướng, chúng ta cũng sẽ thấy rằng sự giống nhau giữa hai phép đo sẽ thay đổi theo hướng. Bằng cách hiểu các giá trị dữ liệu thay đổi như thế nào theo khoảng cách và hướng, ta có thể nội suy các giá trị tại các vị trí chưa được lấy mẫu trong toàn bộ khu vực nghiên cứu của mình.

### *Nguồn gốc của địa thống kê*

Nguồn gốc của thống kê địa lý chỉ được tìm thấy trong ngành khai thác mỏ. D.G. Krige, một kỹ sư khai thác mỏ người Nam Phi và H.S. Sichel, một nhà thống kê đã phát triển phương pháp ước tính này vào đầu những năm 1950, khi các số liệu thống kê “cổ điển” được cho là không phù hợp để ước tính trữ lượng quặng rải rác.

Georges Matheron, một kỹ sư người Pháp, đã phát triển các khái niệm đổi mới của Krige và chính thức hóa chúng trong một khuôn khổ duy nhất với Lý thuyết về các biến khu vực hóa của ông (Matheron, 1970, bằng tiếng Pháp). Thuật ngữ "kriging" được đặt ra để ghi nhận D.G. Krige.

Mặc dù ban đầu được phát triển để giải các bài toán ước tính trữ lượng quặng, nhưng sự ra đời của máy tính tốc độ cao vào những năm 1970 đã tạo điều kiện cho sự lan rộng của địa thống kê từ khai thác mỏ và địa chất sang một loạt các ngành khác (bao gồm sinh học, động vật học, sinh thái học, tâm lý học và y học). Mãi cho đến giữa đến cuối những năm 1980, các kỹ thuật địa thống kê mới được sử dụng ở bất kỳ mức độ nào trong ngành dầu khí và mức độ phổ biến của nó đã tăng lên hàng năm kể từ đó.

Trong ngành công nghiệp dầu khí, địa thống kê được sử dụng chủ yếu để mô tả một cách định lượng đặc tính bể chứa. Địa thống kê cung cấp khung khái niệm để xác định kiến trúc bên trong của bể chứa bằng cách kết hợp kiến thức chủ quan và khách quan về bể chứa để tạo ra các mô hình chi tiết về các đặc tính của nó.

### *Vai trò của địa thống kê trong đặc tính bể chứa*

Chi phí khổng lồ cho việc phát triển các mỏ ngoài khơi không đồng nhất và mong muốn tăng khả năng phục hồi cuối cùng đã thúc đẩy các công ty dầu khí sử dụng các kỹ thuật mô tả đặc tính bể chứa một cách sáng tạo để xác định quy luật các đặc tính khác nhau được phân bố khắp bể chứa. Địa thống kê là một trong nhiều công nghệ mới thường được tích hợp vào quá trình mô tả đặc tính bể chứa. Nó được sử dụng như một phương tiện tính toán giá trị của các thuộc tính giữa các điểm dữ liệu được đo thực tế (nội suy), do đó tạo ra một lưới các giá trị có thể được sử dụng để tạo bản đồ, mặt cắt ngang và mô hình dòng chảy.

Trong thập kỷ qua, các kỹ thuật địa thống kê, đặc biệt là các kỹ thuật kết hợp dữ liệu địa chấn 3D, đã được chấp nhận như một cách để mô tả các bể chứa dầu khí, với các mô tả bằng số thu được thường được sử dụng làm đầu vào cho các mô phỏng dòng chảy chất

lỏng. Là một thành phần chính của Đặc tính hồ chứa tích hợp, Địa thống kê yêu cầu sự hợp tác giữa các nhà địa chất, nhà địa vật lý, nhà địa hóa học và kỹ sư dầu khí; tạo điều kiện cho mỗi chuyên ngành đóng góp đầy đủ vào quá trình xây dựng mô hình bể chứa.

Cách tiếp cận này khá khác so với trước đây, nơi mà việc hình thức hóa toán học thường được giao cho các kỹ sư làm công tác xây dựng mô hình bể chứa, những người đã quen thuộc hơn với toán học và máy tính. Vì vậy, một phần của triết lý địa thống kê là đảm bảo rằng thực tế địa chất không bị mất đi trong quá trình này (Yarus, 1994).

Quá trình lập mô hình bể chứa liên quan đến việc chuyển đổi các mô hình địa chất định tính thành một mô hình số, thường do một người nào đó không phải là nhà địa chất thực hiện. Nếu mô hình ban đầu là chính xác, thì việc chuyển đổi sẽ không gây ra vấn đề gì. Thật không may, thường thì mô hình số cuối cùng không giống với mô hình ban đầu. Yarus (1994) cho rằng sự khác biệt thường không phải là kết quả của cách giải thích chuyên nghiệp, mà thường dựa trên các lý do thực dụng. Để xây dựng các mô hình bể chứa rất tốn kém cả về mặt thời gian và chi phí, đòi hỏi rất nhiều thời gian của máy tính (và thời gian của con người) và nhiều thao tác với tệp số liệu. Để mô phỏng một bể chứa ở độ phân giải rất tốt là không hợp lý và không thể chấp nhận được về mặt tài chính. Do đó, để giảm thời gian (và chi phí) mô phỏng máy tính, kỹ sư bể chứa chuyển đổi mô hình địa chất dưới dạng mịn và liên tục ban đầu thành một số nút lưới để dễ quản lý và xây dựng hơn.

Có sự phân nhánh để giảm đáng kể số lượng ô trong mô hình bể chứa. Đầu tiên, tính không đồng nhất hoặc phức tạp của địa chất đã được đơn giản hóa quá mức. Nếu bể chứa không quá phức tạp, thì một biểu diễn thô hơn có thể phù hợp; tuy nhiên, kết quả mô phỏng thô thường gây hiểu lầm cho các hồ chứa phức tạp. Để ngăn chặn điều này, các kỹ thuật khớp lịch sử đã được sử dụng để tinh chỉnh mô hình kỹ thuật thô hơn. Độ xấp, độ thấm và các thông số khác được “điều chỉnh” cho đến khi mô phỏng dòng chất lỏng phù hợp với hiệu suất giếng quan sát được hoặc áp suất và tốc độ dòng chảy từ các thử nghiệm sản xuất. Khi bất kỳ điều kiện nào trong số này được đáp ứng, người ta cho rằng mô hình là chính xác. Tuy nhiên, không phải lúc nào cũng như vậy (Yarus, 1994).

Mặc dù mô phỏng bể chứa mang lại một kết luận nhất định cho một nghiên cứu bằng cách cung cấp kế hoạch phát triển và kinh tế, hiệu suất sản xuất trong tương lai thường gây thất vọng. Các dự đoán ban đầu về sản xuất và dự trữ có thể phục hồi thường bị nhầm lẫn. Các giếng riêng lẻ không hoạt động như dự kiến, các kiểu khoan lấp có thể không phù hợp và các phương pháp phục hồi nâng cao thường được yêu cầu.

Địa thống kê cố gắng cải thiện dự đoán bằng cách phát triển một loại mô hình định lượng khác. Mục tiêu là xây dựng một mô hình thực tế hơn về tính không đồng nhất của hồ chứa bằng cách sử dụng các phương pháp không lấy trung bình các thuộc tính quan trọng của hồ chứa. Giống như cách tiếp cận tất định truyền thống, nó bảo tồn dữ liệu “cứng” không thể chối cãi ở những nơi nó được biết đến, đồng thời tôn trọng dữ liệu “mềm” có thể diễn giải ở những nơi mong muốn. Tuy nhiên, không giống như cách tiếp cận tất định, địa thống

kê cung cấp cho các nhà khoa học nhiều kết quả hợp lý. Mức độ khác nhau của các mô hình khác nhau là sự phản ánh của điều chưa biết, hoặc phép đo “sự không chắc chắn”. Yarus (1994) tuyên bố rằng một số kết quả có thể thách thức trí tuệ địa chất phổ biến, và gần như chắc chắn sẽ cung cấp một loạt các kịch bản kinh tế từ lạc quan đến bi quan.

## II. CÁC YẾU TỐ CƠ BẢN CỦA MỘT NGHIÊN CỨU ĐỊA THỐNG KÊ

Nghiên cứu địa thống kê điển hình sẽ bao gồm các bước sau:

1. Thu thập và chuẩn bị dữ liệu (bao gồm kiểm soát chất lượng ban đầu)
2. Tải dữ liệu
3. Phân tích dữ liệu khám phá
4. Phân tích tính liên tục trong không gian
  - a. tính toán mô hình phương sai thực nghiệm
  - b. minh giải và mô hình hóa
5. Tìm kiếm “Thiết kế Vùng lân cận”
6. Xác thực chéo mô hình
7. Nội suy không gian của các đặc tính bề chứa
8. Mô phỏng có điều kiện các đặc tính của bề chứa
9. Đánh giá độ không đảm bảo của mô hình

### *Hạn chế của phương pháp tiếp cận bằng địa thống kê*

Địa hống kê sẽ không bao giờ thay thế được dữ liệu tốt (và không bao giờ có đủ dữ liệu tốt). Nó cũng không phải là một giải pháp thay thế cho việc tiến hành phân tích kỹ lưỡng dữ liệu. Nó không phải là một kỹ thuật ma thuật hay thuốc chữa bách bệnh. Các mô hình phải được giải thích và xác nhận dựa trên tất cả các thông tin có sẵn và liên quan đến các nguyên tắc địa chất bề chứa, vật lý đá và kỹ thuật bề chứa.

Cũng như đề cập trong phần mở đầu, báo cáo này chỉ tập trung về việc xây dựng bản đồ trường địa vật lý nên dưới đây sẽ trình bày về thuật toán kriging

### III. TỔNG QUAN VỀ KRIGING

#### *Giới thiệu*

Bản đồ đường đồng mức bằng tay hoặc bằng máy tính yêu cầu sử dụng một số loại quy trình nội suy. Nói đến dữ liệu bản đồ không thể không nói đến phép chia lưới và phép nội suy, có một số thuật toán cho phép nội suy dựa trên máy tính. Các nhà địa chất vẽ bản đồ bằng tay cần nội suy giữa các điểm dữ liệu (hoặc ngoại suy ngoài các điểm kiểm soát), vẽ các đường viền, làm mịn bản đồ để làm cho nó trông “thật” và điều đó dẫn đến có thể làm sai lệch bản đồ dựa bằng một cách rất vô tình theo kinh nghiệm địa chất (Hohn, 1988). Phần này cung cấp một cái nhìn bao quát về quy trình nội suy sử dụng nhiều máy tính, nằm ở trung tâm của mô hình địa thống kê.

#### *Ước lượng tuyến tính*

*Kriging* là một kỹ thuật thống kê địa lý để ước tính các giá trị thuộc tính tại một *điểm*, trên một *khu vực* hoặc trong một *thể tích*. Nó thường được sử dụng để nội suy các giá trị nút lưới trong các ứng dụng vẽ bản đồ và đường đồng mức. Về lý thuyết, không có quy trình nội suy nào khác có thể tạo ra các ước tính tốt hơn (không chệch, với sai số tối thiểu); mặc dù hiệu quả của kỹ thuật này thực sự phụ thuộc vào việc mô hình hóa chính xác variogram. Độ chính xác của các ước tính kriging được thúc đẩy bởi việc sử dụng các mô hình variogram để thể hiện mối quan hệ tự tương quan giữa các điểm kiểm soát trong tập dữ liệu. Kriging cũng tạo ra một ước tính phương sai cho các giá trị nội suy của nó.

Kỹ thuật này lần đầu tiên được sử dụng để ước tính trữ lượng và cấp quặng vàng ở Nam Phi (do đó là nguồn gốc của thuật ngữ Hiệu ứng *Nugget*), và nó được đặt tên để vinh danh một kỹ sư khai thác mỏ người Nam Phi, Danny Krige. Giá trị và nền tảng toán học được phát triển bởi Georges Matheron, người sau này thành lập *Trung tâm địa thống*, là một phần của *Ecole des Mines* ở Paris, Pháp. (Henley, 1981; Hohn, 1988; Journel, 1989; Isaaks và Srivastava, 1989; Deutsch và Journel, 1992; Wackernagel, 1995).

#### *Tính năng Kriging*

Kriging là một quá trình ước tính có độ chính xác cao:

- giảm thiểu sai số ước lượng (chênh lệch giữa giá trị *đo được* - giá trị *ước tính lại*)
- tôn trọng dữ liệu “cứng”
- không giới thiệu sai lệch ước tính
- không tái tạo sự biến thiên giữa các giếng

- tạo ra một kết quả “làm mịn”; giống như tất cả các bộ nội suy
- là một ước lượng đơn biến; chỉ cần một mô hình hiệp phương sai
- cân điểm không chế theo mô hình không gian (variogram)
- có xu hướng về giá trị *trung bình* khi dữ liệu thưa thớt
- sử dụng mô hình tương quan không gian để xác định trọng số ( $\lambda$ )
- gán trọng số *âm* hoặc *null* cho các điểm kiểm soát bên ngoài phạm vi tương quan của mô hình không gian
- chỉ ra *độ tin cậy toàn cầu tương đối* (*global relative reliability*) của ước tính thông qua *lỗi RMS* (phương sai kriging), như một sản phẩm phụ của quá trình kriging
- có một ma trận tổng quát và dễ dàng định dạng lại, làm cho nó trở thành một kỹ thuật rất linh hoạt để sử dụng nhiều hơn một biến
- loại bỏ bó dữ liệu (declustering) trước khi ước tính

### ***Các loại Kriging***

Có một số thuật toán kriging và mỗi thuật toán được phân biệt bằng cách xác định và sử dụng giá trị trung bình trong quá trình ước tính. Bốn phương pháp được sử dụng phổ biến nhất là:

- ***Simple Kriging*** : *Giá trị trung bình toàn cầu* được biết (hoặc được cung cấp bởi người dùng) và được giữ cố định trên toàn bộ khu vực nội suy.
- ***Kriging thông thường*** : *Giá trị trung bình cục bộ* khác nhau và được ước tính lại dựa trên các điểm kiểm soát trong hình elip lân cận tìm kiếm hiện tại.
- ***Kriging with an External Drift (Kriging với sự trôi bên ngoài)***: Mặc dù phương pháp này sử dụng hai biến, nhưng chỉ cần một mô hình hiệp phương sai và hình *dạng* của bản đồ có liên quan đến thuộc tính 2-D hướng dẫn phép nội suy của thuộc tính chính chỉ được biết tại các vị trí rời rạc. Một ứng dụng điển hình là chuyển đổi thời gian sang độ sâu, trong đó thuộc tính *chính* (chẳng hạn như độ sâu tại các giếng) có được hình dạng của nó từ thuộc tính *thứ cấp*, được gọi là trôi dạt bên ngoài (chẳng hạn như thời gian di chuyển hai chiều được biết đến trên mô hình 2-D lưới).

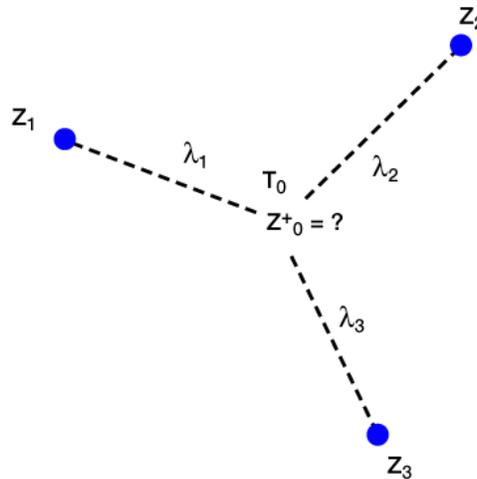
- **Chỉ báo Kriging** : ước tính xác suất của một thuộc tính tại mỗi nút lưới (ví dụ: thạch học, sản lượng). Kỹ thuật này yêu cầu các tham số sau:

- Mã hóa thuộc tính ở dạng nhị phân, là 0 hoặc 1.
- Xác suất trước của cả hai lớp.
- Mô hình hiệp phương sai không gian của biến chỉ báo.

### Quá trình Kriging

Chúng tôi sẽ minh họa quá trình ước lượng bằng một bài toán ví dụ, như trong hình 1.

Các mẫu đã cho ở ( $Z_\alpha$ ), trong đó  $\alpha= 1, 2, 3$  Tìm giá trị có khả năng nhất của biến  $Z$  tại điểm đích (nút lưới:  $Z_0^*$ , [Hình 1](#)). Trong đồ họa này, chúng ta thấy sự sắp xếp hình học của ba điểm dữ liệu  $Z_\alpha$ , vị trí của điểm có giá trị mà chúng ta muốn ước tính  $Z_0^*$  và các trọng số chưa biết,  $\lambda_\alpha$ .



Hình 1: Sắp xếp ba điểm dữ liệu

Coi  $Z_0^*$  là tổ hợp tuyến tính của dữ liệu  $Z_\alpha$

- $Z_0^* = \lambda_0 + \sum \lambda_\alpha Z_\alpha$
- Ở đâu:  $\sum \lambda_\alpha = 1$  và  $\lambda_0 = m_z - \sum \lambda_\alpha$
- Xác định  $\lambda_\alpha$  sao cho:

- $Z_0^*$  không chệch:  $E[Z_0^* - Z_\alpha] = 0$
- $Z_0^*$  có sai số bình phương trung bình nhỏ nhất (MSE- mean square error ())
- $E[Z_0^* - Z_\alpha]^2$  là tối thiểu

Nhớ lại rằng giá trị chưa biết  $Z_0^*$  được ước tính bằng tổ hợp tuyến tính của  $n$  điểm dữ liệu cộng với tham số dịch chuyển  $\lambda_0$ :

$$Z_0^* = \lambda_0 + \sum \lambda_\alpha Z_\alpha \quad (1)$$

Bằng cách biến đổi phương trình trên thành một tập hợp các phương trình bình thường tuyến tính, chúng tôi giải các phương trình sau để thu được các trọng số  $\lambda_\alpha$ . Tập hợp các phương trình tuyến tính có dạng sau:

$$\sum \lambda_j \underline{C}(x_\alpha, x_j) - \mu = \underline{c}(x_\alpha, x_0) \text{ với mọi } j = 1, n \quad (2)$$

hoặc trong ký hiệu tốc ký ma trận:

$$\underline{C} \underline{\Lambda} = \underline{c} \quad (3)$$

Cả ba số hạng đều là ma trận trong đó:

- $\underline{C}(x_\alpha, x_j)$  biểu thị hiệp phương sai (covariance) giữa các điểm mẫu  $x_\alpha$  và  $x_j$
- $\underline{c}(x_\alpha, x_0)$  biểu thị hiệp phương sai giữa một mẫu ở  $x_\alpha$  và điểm đích  $x_0$ ; điểm ước tính
- là các trọng số chưa biết,  $\lambda_j$
- là một hệ số nhân Lagrange chuyển đổi bài toán cực tiểu hóa có ràng buộc thành bài toán cực tiểu hóa không ràng buộc.

Xác định ma trận của các trọng số chưa biết bằng cách giải phương trình ma trận  $\underline{\Lambda}$  như sau:

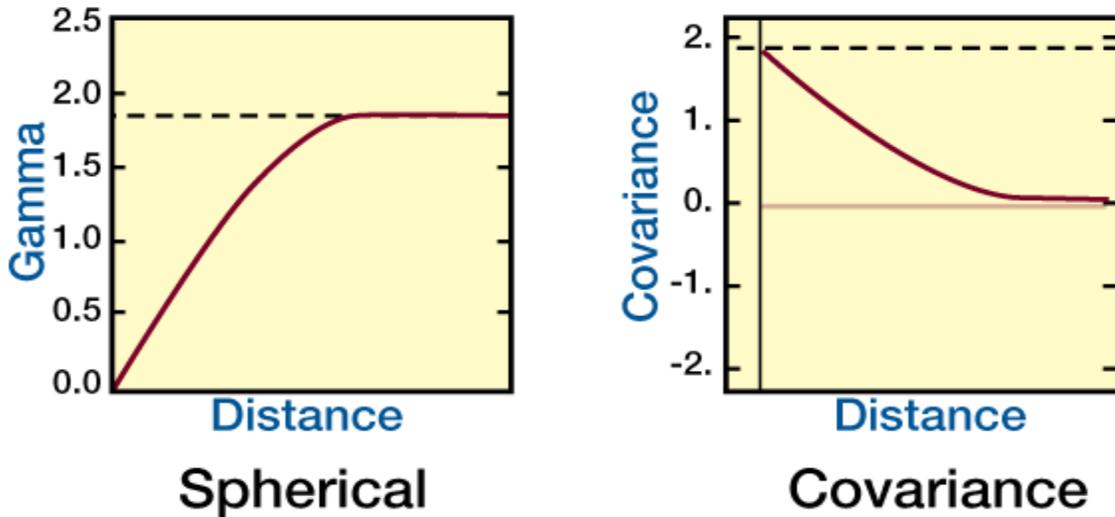
$$\underline{C} \underline{\Lambda} = \underline{c} \quad (4)$$

Với

$$\underline{\Lambda} = \underline{C}^{-1} \underline{c} \quad (5)$$

Lưu ý rằng phương trình 3 được viết dưới dạng các giá trị hiệp phương sai, tuy nhiên, mô hình biểu đồ biến thiên hoặc biểu đồ tương quan đã được lập, chứ không phải hiệp phương sai. Chúng ta sử dụng các giá trị hiệp phương sai vì nó hiệu quả hơn về mặt tính toán.

Hiệp phương sai bằng ngưỡng cửa sổ trừ đi biểu đồ biến thiên ( [Hình 2](#) ) :



Hình 2: *Mối quan hệ giữa biểu đồ biến thiên hình cầu và tương đương hiệp phương sai của nó* ):

$$C_{(h)} = \sigma^2 (\text{ngưỡng cửa}) - \gamma_{(h)} \quad (6)$$

### ***Phương sai Kriging***

Ngoài việc ước tính giá trị của một biến tại một vị trí chưa được lấy mẫu, kỹ thuật kriging còn cung cấp ước tính về lỗi có thể xảy ra (ở dạng phương sai lỗi) tại mọi nút lưới. Các ước tính lỗi này có thể được thể hiện trên bản đồ, để đưa ra đánh giá trực tiếp về độ tin cậy của các đường đồng mức trên bề mặt. Bởi vì các phương sai cắt bỏ được xác định độc lập với các giá trị dữ liệu, nên *sai số cắt bỏ* không phải là thước đo độ tin cậy cục bộ (Deutsch và Journel, 1992). Đừng cố gắng sử dụng *độ lệch chuẩn kriging* như thống kê độ lệch chuẩn cổ điển thực sự.

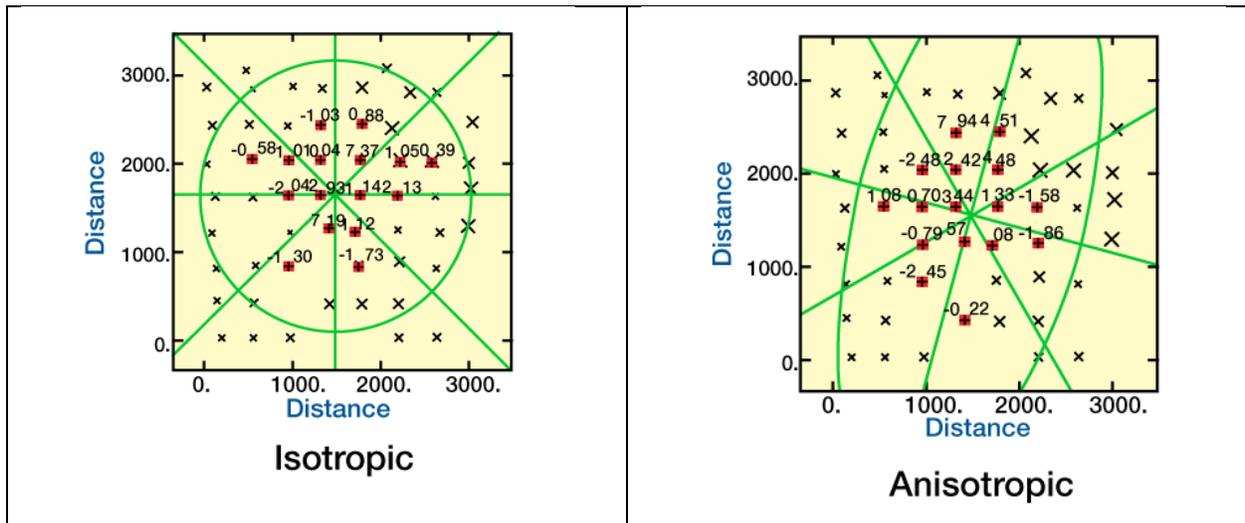
Phương trình phương sai kriging là:

$$\sigma^2_k = C(x_0, x_0) \sum \lambda_{\alpha i} (x_{\alpha}, x_0) - \mu \quad (7)$$

Tìm kiếm tiêu chí vùng lân cận

Tất cả các thuật toán nội suy đều yêu cầu tiêu chí lựa chọn dữ liệu tốt, được chỉ định bởi *vùng lân cận tìm kiếm*. Biến đổi mô hình đóng vai trò kiểm soát phạm vi của vùng lân cận. Phạm vi variogram xác định kích thước tối đa của vùng lân cận mà từ đó các điểm kiểm soát sẽ được chọn để ước tính một nút lưới, nhằm tận dụng mối tương quan thống kê giữa các quan sát.

Một quy trình thống kê địa lý điển hình có thể nội suy các giá trị cho một vị trí cụ thể bằng cách sử dụng các giá trị lân cận gần nhất được tính trọng số theo khoảng cách và mức độ tự tương quan hiện có cho khoảng cách đó (như được xác định bởi mô hình variogram). Việc tìm kiếm vùng lân cận sẽ được giới hạn ở một số lượng hàng xóm gần nhất được chỉ định và cũng có thể bị giới hạn ở một hướng cụ thể.



Hình 3: Sự phân bố số liệu a) đẳng hướng và b) bất đẳng hướng.

Các tham số vùng lân cận tìm kiếm bao gồm:

- Bán kính tìm kiếm
- Hình dạng lân cận (đẳng hướng hoặc bất đẳng hướng): Hình 3a và 3b
- Số lĩnh vực (4 hoặc 8 là phổ biến)

- Số lượng điểm dữ liệu trên mỗi lĩnh vực
  - Vùng lân cận duy nhất sử dụng tất cả các điểm dữ liệu (giới hạn thực tế là 100)
  - Vùng lân cận di chuyển sử dụng một số điểm giới hạn cho mỗi khu vực (ví dụ: 4)
- Phương vị của trục chính bất đẳng hướng

Trong hình 3, lưu ý hình dạng elip của vùng lân cận tìm kiếm bất đẳng hướng và hình tròn cho vùng lân cận đẳng hướng. Cả hai vùng lân cận được chia thành các quãng tám, với tối đa hai điểm dữ liệu cho mỗi khu vực.

Đồ họa này cho thấy bán kính của vùng lân cận bất đẳng hướng là: trục phụ = 1000 mét và trục chính = 4000 mét, được căn chỉnh tại N15E. Mô hình đẳng hướng có chiều dài bán kính 1500 mét. Trung tâm của vùng lân cận là nút lưới mục tiêu để ước tính. Có 55 điểm lấy mẫu (x) trong khu vực nghiên cứu. Trọng số được hiển thị cho các điểm kiểm soát dữ liệu được sử dụng để ước tính tại điểm mục tiêu.

### ***Cần nhắc thực tế trong việc thiết kế vùng lân cận tìm kiếm***

- Căn chỉnh trục tìm kiếm với hướng bất đẳng hướng tối đa.
- Bán kính tìm kiếm (nếu dị hướng) phải là  $\leq$  các khoảng tương quan.
- Mỗi góc phần tư phải có đủ điểm ( $\sim 4$ ) để tránh sai lệch lấy mẫu theo hướng.
- Các yêu cầu về thời gian và bộ nhớ của CPU tăng nhanh như một hàm của số điểm dữ liệu trong vùng lân cận.
- *Về lý thuyết*, nhiều dữ liệu hơn trong hệ thống phân loại làm giảm lỗi bình phương trung bình.
- *Trong thực tế*, hiệp phương sai ít được biết đến với khoảng cách vượt quá khoảng  $\frac{1}{2}$  đến  $\frac{2}{3}$  kích thước của trường. Bao gồm các điểm ở xa hơn thực sự có thể làm tăng lỗi.
- Công cụ ước tính kriging được xây dựng từ dữ liệu trong vùng lân cận tìm kiếm tập trung tại vị trí mục tiêu  $Z_o^*$ .

### ***Cần nhắc thực tế: Vùng lân cận tìm kiếm duy nhất so với di chuyển***

- Trong *dịch chuyển* để xác định vùng lân cận, một hệ phương trình kriging đơn giản (SK) hoặc thông thường (OK) mới được giải tại mỗi nút lưới.
- Trong lần tìm kiếm cận *duy nhất* sử dụng tất cả dữ liệu, do đó có khái niệm tìm kiếm duy nhất, ma trận **C**-phía bên trái của ma trận kriging, chỉ được giải quyết một lần và được sử dụng tại mỗi nút lưới.
- Nếu có đủ giếng để tính toán OK, thì khu vực lân cận đang di chuyển sẽ thích hợp hơn khu vực lân cận duy nhất.
- Các vùng lân cận duy nhất có xu hướng ngăn cản sự thay đổi đột ngột các giá trị và số điểm của các điểm dữ liệu.
- Các vùng lân cận duy nhất làm mịn dữ liệu hơn các vùng lân cận chuyển động.

### ***Cần nhắc thực tế: Kriging thông thường (OK) so với Kriging đơn giản (SK)***

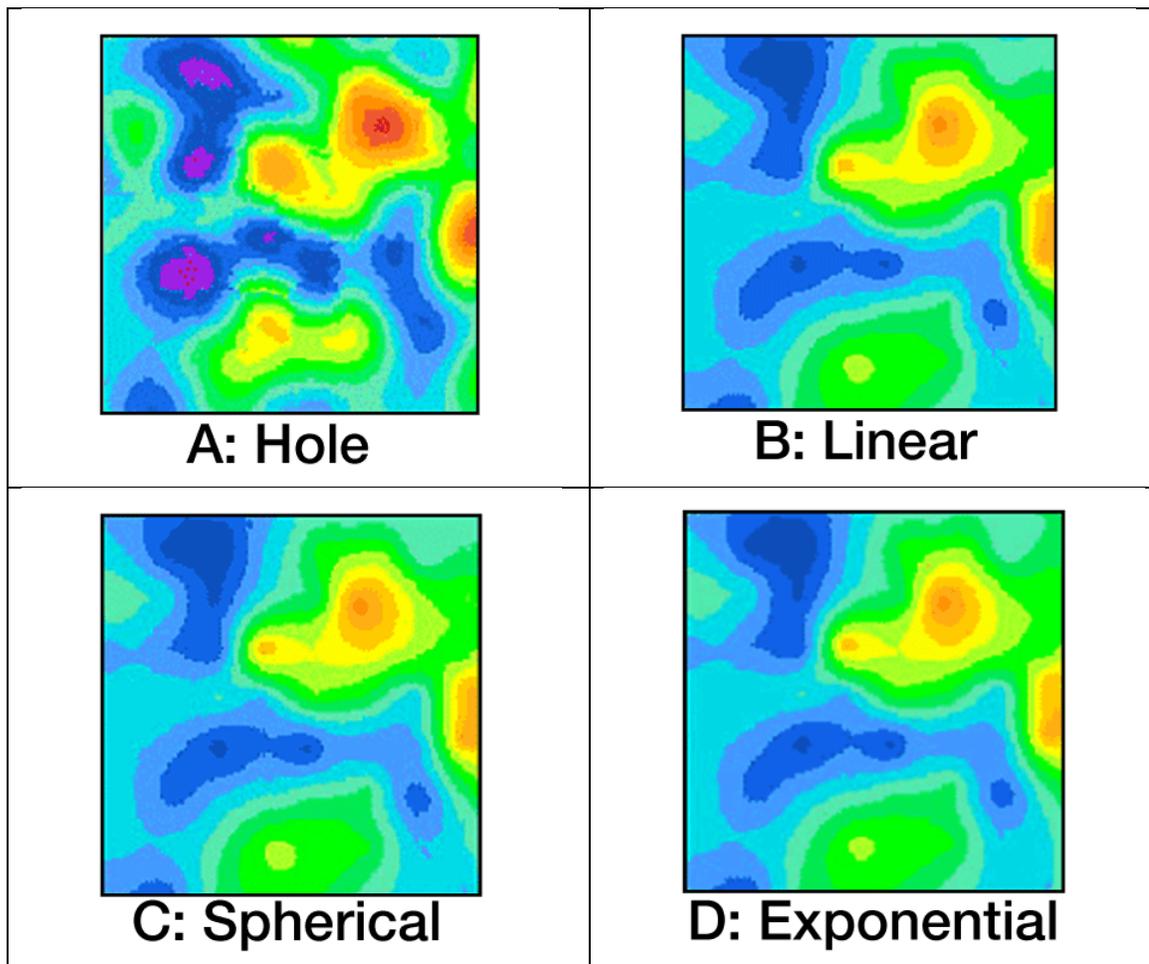
- Kriging đơn giản không thích ứng với xu hướng địa phương; thay vào đó, nó dựa trên một giá trị trung bình toàn cầu, không đổi.
- Kriging thông thường sử dụng giá trị trung bình cục bộ ( $m_z$ ), tương đương với việc ước tính lại  $m_z$  tại mỗi nút lưới từ dữ liệu trong vùng lân cận tìm kiếm.
- Khi tất cả các điểm dữ liệu được sử dụng (láng giếng duy nhất), **OK** and **SK** mang lại kết quả tương tự nhau.
- Nếu chỉ có một vài điểm dữ liệu có sẵn trong vùng lân cận tìm kiếm cục bộ, **OK** có thể tạo ra các trọng số giả do ràng buộc rằng các trọng số phải có tổng bằng 1.
- Nếu các giếng đã biết cung cấp phương pháp lấy mẫu sai lệch, thì tốt hơn là áp dụng **mz của riêng bạn** với phương pháp **SK** hơn là sử dụng với phương pháp **OK**.

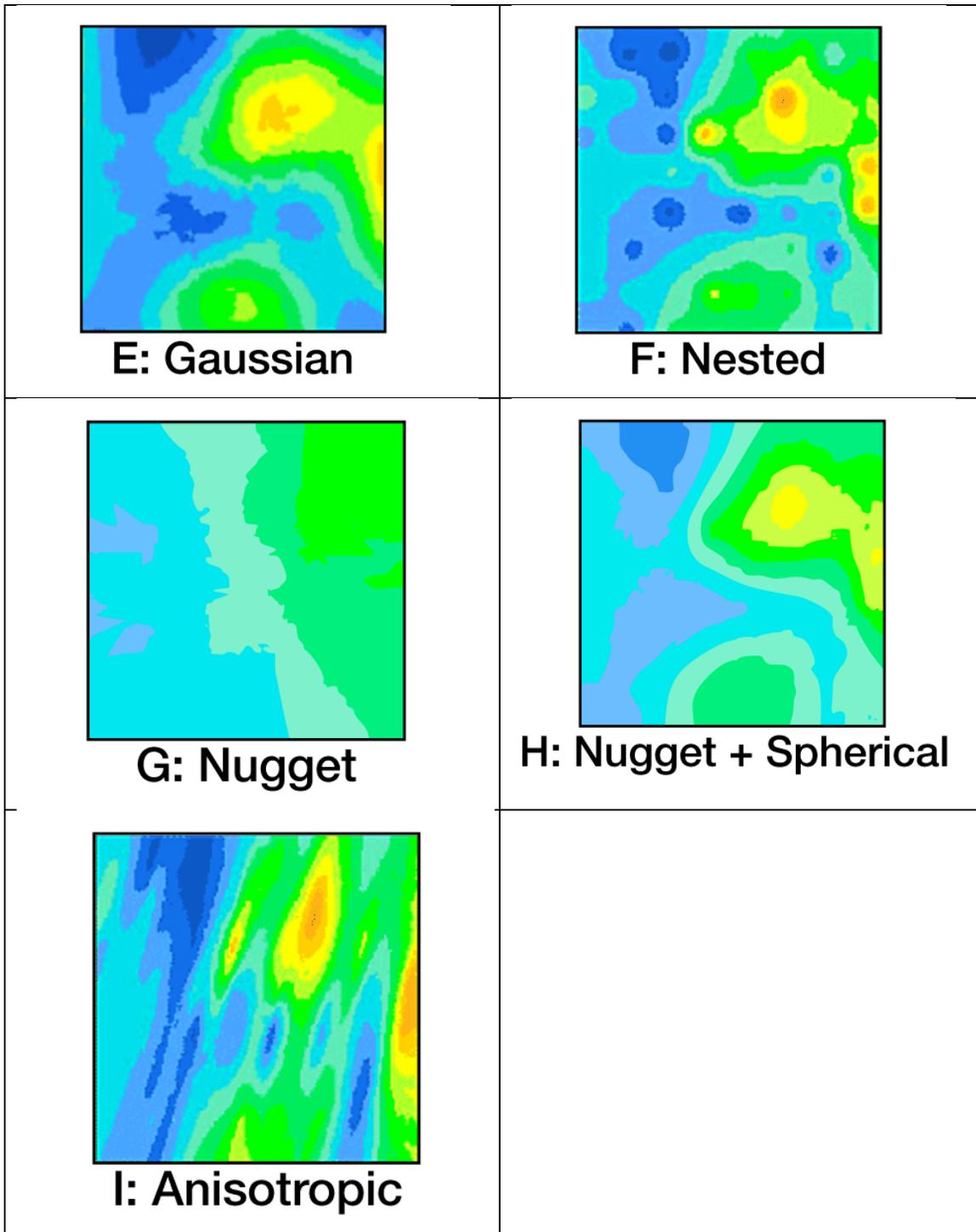
### ***Ảnh hưởng của các thông số Variogram đối với Kriging***

Kriging áp dụng các hàm trọng số theo mô hình toán học của biểu đồ biến đổi. Các ước tính Kriging thu được là ước tính tuyến tính không thiên vị tốt nhất của bề mặt, với điều kiện là bề mặt đó đứng yên và dạng chính xác của biểu đồ biến thiên đã được xác định. Khi hình dạng của biểu đồ biến đổi mô hình thay đổi, thì kết quả cắt giảm cũng vậy.

- Định cỡ lại biểu đồ biến đổi hoặc biểu đồ tương quan (để tạo ngưỡng lớn hơn hoặc nhỏ hơn):

- Không ảnh hưởng đến ước tính kriging
  - Thay đổi phương sai kriging
- Tăng thành phần nugget:
    - Hoạt động như một thuật ngữ làm mịn trong quá trình mài (làm cho trọng lượng giống nhau hơn)
  - Tăng phạm vi có xu hướng tăng ảnh hưởng của nhiều điểm dữ liệu khoảng cách hơn và dẫn đến bản đồ trơn hơn.
  - Hình dạng của biểu đồ biến thiên hoặc biểu đồ tương quan gần gốc ảnh hưởng đến tính liên tục của quá trình nội suy (ví dụ: độ dốc càng thoải thoải thì nội suy càng mịn). Xem Hình 4





**Hình 4:** Kết quả Kriging từ một bộ dữ liệu chung, dựa trên các mô hình biến đổi khác nhau.

Trong hình 4 từ A đến I sử dụng thiết kế vùng lân cận đẳng hướng được minh họa trong hình 3. Mô hình lưới ( $F$ ) đã sử dụng hai phương sai hình cầu, với phạm vi ngắn = 1000 mét và phạm vi dài là 10.000 mét. Ngoài ra còn có các mô hình sử dụng các lưới khác nhau. Mô hình dị hướng ( $I$ ) đã sử dụng thiết kế vùng lân cận dị hướng như trong Hình 3. Các trục phụ của mô hình variogram = 1000 mét, với trục chính = 5000 mét (tỷ lệ bất đẳng hướng 5:1), được quay thành N15E. Thang màu tương đương với tất cả các hình. Màu tím là 5% độ xấp và màu đỏ là 13%. Tất cả những hình minh họa này được tạo bằng cách sử dụng cùng một bộ dữ liệu đầu vào.

### ***Ưu điểm của Kriging***

- Kriging là một bộ nội suy chính xác (nếu điểm điều khiển trùng với một nút lưới).
- Phương sai Kriging:
  - Chỉ số tương đối về độ tin cậy của ước lượng ở các vùng khác nhau.
  - Chỉ báo tốt về hình học dữ liệu.
  - Nugget (hoặc ngưỡng cửa) nhỏ hơn sẽ tạo ra phương sai nhỏ hơn.
  - Giảm thiểu Sai số Bình phương Trung bình.
  - Có thể sử dụng mô hình không gian để điều khiển quá trình nội suy.
  - Một kỹ thuật hiệu quả (tức là, những thay đổi nhỏ trong các thông số xử lý vi sinh tương đương với những thay đổi nhỏ trong kết quả).

### ***Nhược điểm của Kriging***

Kriging có xu hướng tạo ra những hình ảnh trơn hơn thực tế (giống như tất cả các kỹ thuật nội suy). Khi làm như vậy, làm mất hay giảm những biến thiên lớn trong khoảng cách ngắn, (đánh mất các giá trị cao hoặc thấp). Nó cũng yêu cầu đặc điểm kỹ thuật của mô hình hiệp phương sai không gian, có thể khó suy luận từ dữ liệu thưa thớt.

Kriging tiêu tốn nhiều thời gian tính toán hơn so với các kỹ thuật chia lưới thông thường, đòi hỏi phải giải nhiều phương trình đồng thời cho mỗi nút lưới được ước tính. Các quy trình sơ bộ để tạo ra hiệp phương sai cũng như thiết kế các vùng lân cận đòi hỏi nhiều nỗ lực. Do đó, kriging có thể không được thực hiện thường xuyên; thay vào đó, nó được sử

dụng tốt nhất cho các dự án có thể chứng minh nhu cầu ước tính chất lượng cao nhất của bề mặt kết cấu (hoặc thuộc tính bề chứa khác) và được hỗ trợ bởi nhiều dữ liệu tốt.

#### IV. XÁC NHẬN CHÉO

Xác thực chéo là một quá trình kiểm tra tính tương thích giữa một bộ dữ liệu, mô hình không gian và thiết kế vùng lân cận. Trong xác thực chéo, mỗi điểm trong mô hình không gian được xóa riêng lẻ khỏi mô hình và sau đó giá trị của nó được ước tính bằng mô hình hiệp phương sai. Bằng cách này, có thể so sánh giá trị *ước tính* với giá trị *thực tế*.

Quy trình gồm các bước sau:

1. Lần lượt xem xét từng điểm kiểm soát.
2. Tạm thời loại bỏ từng điểm kiểm soát khỏi tập dữ liệu.
3. Ước tính lại từng điểm từ dữ liệu xung quanh bằng mô hình hiệp phương sai.
4. So sánh các giá trị ước tính,  $Z_{est}$ , với giá trị đúng,  $Z_{đúng}$ . Điều này cũng cung cấp một lỗi ước tính lại (phương sai kriging cũng được tính toán cùng một lúc):

$$LAI = Z_{est} - Z_{đúng}$$

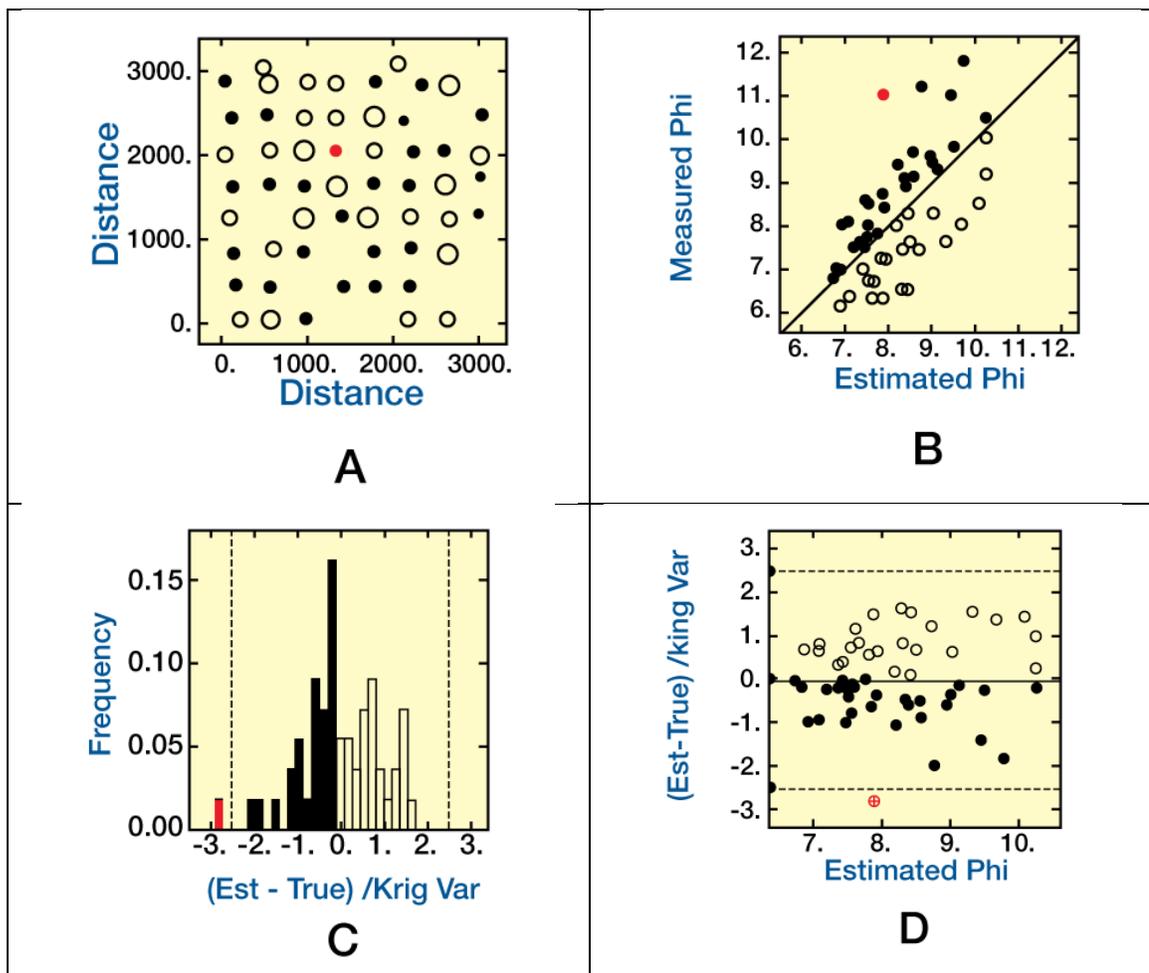
5. Tính sai số chuẩn hóa:

$$SE = RE / \sigma_{krig}$$

- Lý tưởng nhất là nó phải có trung bình bằng 0 và phương sai bằng 1.
  - Tử số bị ảnh hưởng bởi phạm vi (range)
  - Mẫu số bị ảnh hưởng bởi ngưỡng (sill)
6. Trung bình các lỗi cho một số lượng lớn các điểm mục tiêu để đạt được:
    - Lỗi trung bình
    - Sai số trung bình
    - Sai số bình phương trung bình

- Sai số bình phương trung bình chuẩn hóa (Mean squared standardized error)
7. Phân bố lỗi (trong chế độ xem bản đồ) có thể cung cấp tiêu chí hữu ích cho:
- Chọn vùng tìm kiếm
  - Lựa chọn mô hình hiệp phương sai
8. Bất kỳ điểm dữ liệu nào có Lỗi chuẩn hóa tuyệt đối  $\geq 2.5$  được coi là *ngoại lệ*, dựa trên thực tế rằng điểm dữ liệu nằm ngoài giới hạn tin cậy 95% của phân phối chuẩn.

***Sự hữu ích của việc sử dụng các ô xác thực chéo***

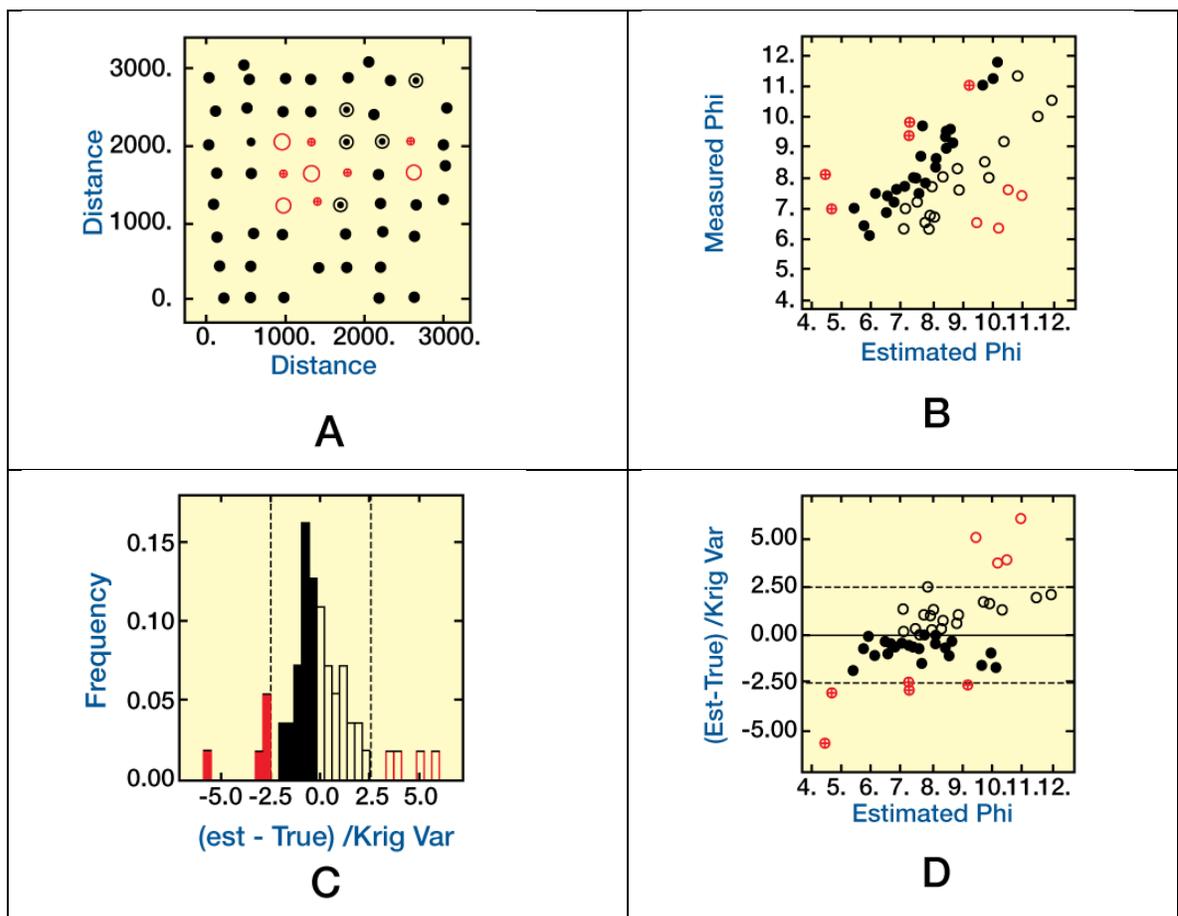


Hình 5: ( *Biểu đồ xác thực chéo* ) hiển thị đầu ra từ thử nghiệm xác thực chéo.

Hình 5a hiển thị chế độ xem bản đồ về mức độ của lỗi ước tính lại (RE). Các vòng kết nối mở vượt quá ước tính; các vòng tròn đặc nằm dưới ước tính. Vòng tròn liền màu đỏ nằm ngoài 2,5 độ lệch chuẩn so với giá trị trung bình = 0. Ngoài ra, hãy tìm kiếm sự trộn lẫn của RE như một dấu hiệu bổ sung về xu hướng. Một mô hình tốt có khả năng ước tính cao hơn hoặc thấp hơn bất kỳ vị trí nào.

Hình 5b là biểu đồ chéo của thuộc tính độ xấp đo được tại các giếng so với độ xấp được ước tính lại tại các vị trí giếng trong quá trình kiểm tra xác nhận chéo. Một lần nữa, các vòng kết nối mở đang được ước tính. Vòng tròn liền màu đỏ là mẫu trong Hình 5a.

Hai biểu đồ quan trọng nhất nằm trong Hình 5c và 5d vì chúng giúp xác định độ chệch của mô hình. Nếu biểu đồ sai số chuẩn hóa (SE) trong Hình 5c bị lệch, nếu hoặc có mối tương quan giữa SE và các giá trị ước tính trong Hình 5d, thì mô hình bị sai lệch. Đó không phải là trường hợp trong ví dụ này; tuy nhiên,



Hình 6: ( Kết quả xác thực chéo từ một mô hình sai lệch )

Trong Hình 6a, các giá trị RE ước tính vượt mức được tập hợp ở trung tâm bản đồ. Biểu đồ của SE (Hình 6c) hơi lệch so với ước tính. Cuối cùng, có một mối tương quan tích cực

giữa SE và độ xấp ước tính. Những điều này cho thấy thiết kế mô hình kém, thiết kế vùng lân cận kém hoặc cả hai.

## **V. KẾT LUẬN và KIẾN NGHỊ**

Đây chỉ mới là nghiên cứu bước đầu của địa thống kê trong việc xây dựng bản đồ. Vấn đề này tại Việt Nam hiện nay còn đang ít được quan tâm. Trong tương lai, khi tiến hành số hóa và xây dựng các bản đồ thuộc tính khu vực nghiên cứu đòi hỏi độ chính xác cao hơn thì vấn đề này cần được xem xét và cân nhắc kỹ lưỡng.