

# CHUẨN HÓA CƠ SỞ DỮ LIỆU VÀ XÂY DỰNG PHẦN MỀM QUẢN LÝ, PHÂN TÍCH CƠ SỞ DỮ LIỆU TÀI NGUYÊN NƯỚC VIỆT NAM

NCS. Hoàng Thanh Tùng <sup>1)</sup>, TS. Ngô Lê An <sup>2)</sup>, SV. Trịnh Quang Toàn <sup>3)</sup>

- 1) Khoa Thủy văn và Tài Nguyên nước, Đại học Thủy lợi  
175 Tây Sơn, Đống Đa, Hà Nội; Email: httung@wru.edu.vn
- 2) Khoa Thủy văn và Tài Nguyên nước, Đại học Thủy lợi  
175 Tây Sơn, Đống Đa, Hà Nội; Email: nlan@wru.edu.vn
- 3) Lớp 47V, Đại học Thủy lợi, Email: mark09870@gmail.com

## Tóm tắt

Đề tài “Chuẩn hóa cơ sở dữ liệu và xây dựng phần mềm quản lý, phân tích cơ sở dữ liệu tài nguyên nước Việt Nam” đi giải quyết các vấn đề khó khăn và những hạn chế hiện nay trong quản lý dữ liệu của ngành tài nguyên nước Việt Nam. Những hạn chế này bao gồm: i) số liệu được lưu trữ không đồng bộ, với nhiều định dạng khác nhau, ii) khuôn thức lưu trữ cũng không thống nhất, không nhất quán về bố cục, iii) rất khó tập hợp dữ liệu về một dạng chuẩn, và quản lý dữ liệu theo hệ thống, iv) chưa có cơ chế đánh giá rõ ràng về chất lượng của số liệu trong cơ sở dữ liệu, v) không có tính kết nối giữa hệ thống các trạm đo đạc, về số liệu và vị trí địa lý giữa các trạm. Với mục đích như vậy, nghiên cứu đã xây dựng được một hệ thống các mô đun bao gồm mô đun chuẩn hóa dữ liệu kết hợp với phần mềm quản lý, phân tích cơ sở dữ liệu cho lĩnh vực “Tài Nguyên Nước Việt Nam” với các chức năng nhằm giải quyết những hạn chế vừa nêu trên đồng thời hỗ trợ đắc lực cho việc tính toán, phân tích thủy văn cũng như kết nối với các mô hình tính toán thủy văn, thủy lực thông dụng hiện nay.

**Các từ khóa:** Chuẩn hóa, Tài nguyên nước, mô đun, cơ sở dữ liệu, định dạng, quản lý

## 1. MỞ ĐẦU

Ngày nay, với sự phát triển nhanh chóng của khoa học kỹ thuật, con người đã đạt được nhiều thành tựu trên mọi lĩnh vực góp phần không nhỏ cải thiện đời sống của nhân loại. Tuy nhiên con người cũng đang phải đối mặt với một khó khăn mới về số lượng thông tin quá lớn, sự bùng nổ về số lượng dữ liệu, dẫn đến việc chọn lọc thông tin, tìm kiếm dữ liệu là một vấn đề hết sức khó khăn. Để giải quyết vấn đề này, con người đã sử dụng “cơ sở dữ liệu” hay còn gọi là tập hợp thông tin có cấu trúc, để quản lý lưu trữ dữ liệu, giúp con người quản lý, chọn lọc và tìm kiếm thông tin nhanh nhất có thể. Hiện nay, trên thế giới đã có khá nhiều các hệ quản trị cơ sở dữ liệu hay các phần mềm lưu trữ, quản lý phân tích dữ liệu khác nhau, và tập trung vào các dữ liệu về đời sống xã hội, kinh tế, chính trị, văn hóa, và một số ngành khoa học. Tuy nhiên, việc lưu trữ và quản lý tài liệu Tài nguyên nước nói chung hay dữ liệu Thủy văn nói riêng ở Việt Nam vẫn chưa được phát triển do một số nguyên nhân sau: i) số liệu được lưu trữ không đồng bộ, với nhiều định dạng khác nhau, ii) khuôn thức lưu trữ cũng không thống nhất, không nhất quán về bố cục, iii) rất khó tập hợp dữ liệu về một dạng chuẩn, và quản lý dữ liệu theo hệ thống, iv) chưa có đánh giá rõ ràng về chất lượng của số liệu, v) không có tính kết nối giữa hệ thống các trạm đo đạc, về số liệu và vị trí địa lý giữa các trạm.

Chính vì vậy, đề tài đã thực hiện nghiên cứu về cách thức chuẩn hóa dữ liệu, nhằm tạo ra một dạng thống nhất để lưu trữ và quản lý dữ liệu mang tính hệ thống. Tiến hành xây dựng phần mềm quản lý, phân tích cơ sở dữ liệu cho lĩnh vực “Tài Nguyên Nước Việt Nam” để giải quyết những nguyên nhân và hạn chế trên, tiến tới một hệ thống dữ liệu ưu việt và thuận tiện nhất cho người sử dụng.

## 2. PHƯƠNG PHÁP NGHIÊN CỨU

Để xây dựng được một hệ thống các môđun bao gồm môđun chuẩn hóa dữ liệu kết hợp với phần mềm quản lý, phân tích cơ sở dữ liệu cho lĩnh vực “Tài Nguyên Nước Việt Nam”, chúng tôi đã tiến hành thực hiện 5 công việc chính sau:

- 1) Nghiên cứu áp dụng phần mềm cơ sở dữ liệu tài nguyên nước có tên là “Water Resources Database” của Hoa Kỳ vào hỗ trợ chuẩn hoá số liệu TNN Việt Nam, đây là phần mềm có chức năng chuẩn hoá số liệu mạnh và mang tính hệ thống..
- 2) Nghiên cứu xây dựng form nền hỗ trợ nhập số liệu cho tất cả các trạm đo đạc KTTV và môi trường của Việt Nam với mã trạm được tổ chức Khí tượng Thế giới - WMO quy định.
- 3) Tiến hành lập trình giao diện, mang tính khoa học nhằm quản lý dữ liệu thuận tiện và đơn giản nhất.
- 4) Lập trình nhằm tạo ra tính kết nối giữa phần mềm “Water Resources Database” với phần mềm “quản lý, phân tích cơ sở dữ liệu cho lĩnh vực Tài Nguyên Nước Việt Nam”, liên kết dữ liệu đã được chuẩn hóa, quản lý dữ liệu dưới dạng số và dạng bản đồ.
- 5) Lập trình tạo ra các phương pháp phân tích đơn giản, và kết nối với các công cụ tính toán cơ bản trong khoa học Khí tượng - Thủy văn với các dữ liệu hiện có.

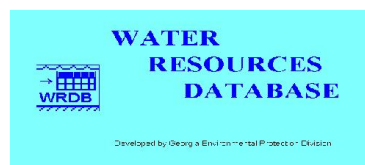
Phần sau đây giới thiệu tóm tắt công cụ chính sử dụng trong nghiên cứu:

#### ***Giới thiệu tóm tắt ngôn ngữ lập trình C#***

C# là ngôn ngữ lập trình thuộc trong hệ .NET, một ngôn ngữ hỗ trợ lập trình hướng đối tượng khá phổ biến hiện nay. C# được phát triển tại Microsoft, do đội nghiên cứu do Hejlsberg và Scott Wiltamuth lãnh đạo. C# không chỉ là ngôn ngữ hướng đối tượng, mà đó còn là ngôn ngữ lập trình trực quan và điều khiển sự kiện. Hơn nữa đây là ngôn ngữ cho phép giao tiếp với các ngôn ngữ máy khác và các chương trình cho phép mọi người truy cập qua bất cứ thiết bị nào. Với công cụ lập trình này, chúng tôi đã tiến hành xây dựng phần mềm “quản lý, phân tích cơ sở dữ liệu cho lĩnh vực Tài Nguyên Nước Việt Nam” bằng ngôn ngữ C# trên nền của WINDOWS FORM thông qua bộ lập trình Microsoft Visual Studio 2005. Phần mềm này đã tích hợp các modun lập trình mà chúng tôi đã thực hiện với phần mềm WRDB - phần mềm chuẩn hoá dữ liệu để tạo thành một phần mềm quản lý cơ sở dữ liệu hoàn chỉnh cả dữ liệu KTTV, môi trường, với số liệu bản đồ.

#### ***Giới thiệu tóm tắt WRDB***

Phần mềm WRDB được Tổ chức Bảo vệ Môi trường Georgia (EPD) xây dựng và phát triển liên tục từ năm 1993 với mục đích quản lý thống nhất một khối lượng lớn những dữ liệu khác nhau về khí tượng, thủy văn và môi trường. Phiên bản mới nhất của phần mềm này là version 5.0.



Phần mềm WRDB cho phép quản lý dữ liệu có hệ thống và cho phép mô tả tất cả các dữ liệu thu thập được một cách hiệu quả và toàn diện.

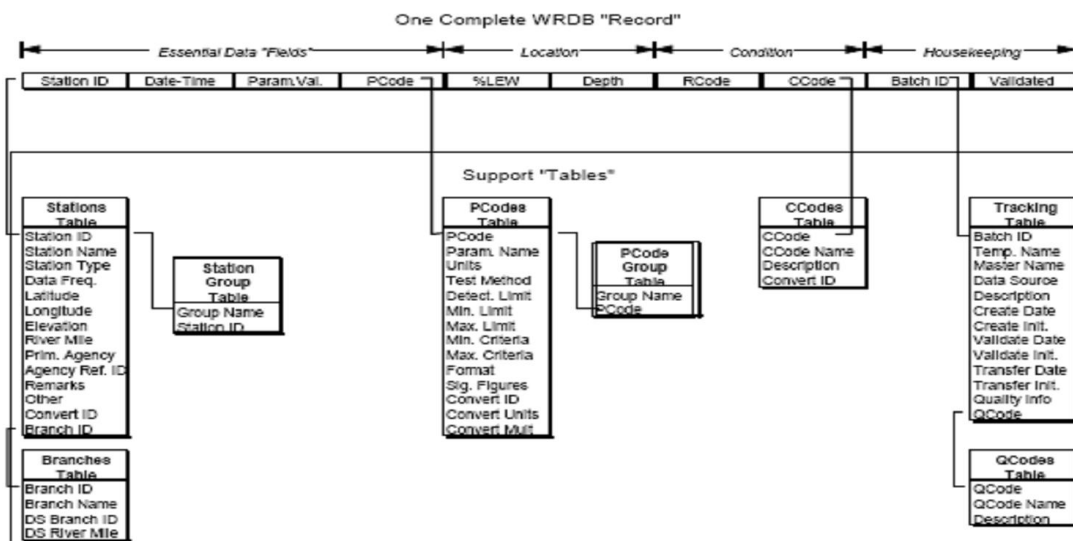
- Cơ sở dữ liệu có thể được quản lý ở hầu hết các định dạng cơ bản như paradox, Access, SQL Server, Oracle, Firebird, Firebird Embedded, MySQL, SQLite vì vậy có thể truy cập nhanh dữ liệu trên PC trong môi trường Windows. Đồng thời, cơ sở dữ liệu này hoàn toàn có thể chuyển sang các định dạng cơ sở dữ liệu hoặc hệ thống quản lý cơ sở dữ liệu như trên.
- Các bảng dữ liệu có thể nhập từ nhiều mẫu khác nhau rất tiện lợi khi hầu hết số liệu thu thập ở Việt Nam ở nhiều dạng bảng biểu khác nhau. Đặc biệt nó có thể được thiết kế để nhập ngay các số liệu quan trắc từ các dataloger đo tự động. Các bảng số liệu của CSDL có thể xuất ra thành những bảng tính luân phiên đa dạng, thành cơ sở dữ liệu hay là định dạng văn bản.
- Toàn bộ các bảng kết hợp có thể được tạo ra, chỉnh sửa và in ấn sử dụng theo kiểu bảng tính đang hiển thị hoặc có thể tiến hành các phép phân tích thông kê đơn giản và hiển thị dưới dạng các biểu đồ thông dụng.

- Cơ sở dữ liệu được phân cấp quản lý nên nó tính bảo mật cao. Đồng thời nó có cơ chế tự kiểm tra và phát hiện ra các trị số khác thường của số liệu có thể sai sót do việc nhập dữ liệu không chính xác v.v... .
- Vị trí các trạm có thể được quản lý trên nền GIS, do có tính kết hợp với phần mềm BASINS 4.

Với những tính năng trên, phần mềm WRDB rất thích hợp và dễ dàng trong việc đưa các dạng dữ liệu khí tượng, thủy văn dưới dạng số từ các nguồn dữ liệu, định dạng khác nhau về cùng một kiểu. Và kết quả truy xuất ra cũng có thể ở nhiều định dạng phổ biến là đầu vào cho nhiều phần mềm tính toán khác.

Cấu trúc của WRDB được mô tả vắn tắt bao gồm các thành phần chính sau và nó được đưa ra trong sơ đồ hình 2.1:

- Các bảng số liệu KTTV - MT (bảng làm việc - Working, bảng Master, và các bảng lựa chọn). Đây chính là phần CSDL chính.
- Các trạm - Stations
- Các nhóm trạm (quản lý các nhóm trạm theo phạm vi lưu vực, tỉnh, ...) – Station groups
- Các nhánh sông - Branches
- Mã các yếu tố đo đạc - Parameter Codes (PCodes)
- Các nhóm mã trạm - Parameter Code Groups
- Phân nhóm các yếu tố đo theo thời gian (ngày, giờ, 6h...) - Compositing Codes (CCodes)
- Báo cáo tóm tắt quản lý việc vào và ra dữ liệu - Database Batch Tracking
- Thông số đánh giá độ tin cậy của số liệu thu thập - Quality Codes (QCodes)



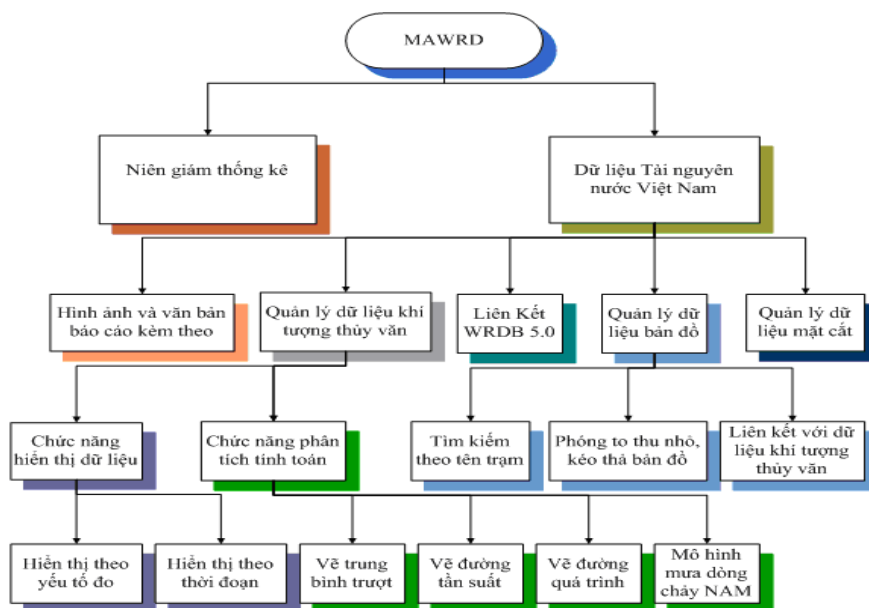
Hình 1. Sơ đồ cấu trúc CSDL trong WRDB

### 3. KẾT QUẢ VÀ THẢO LUẬN

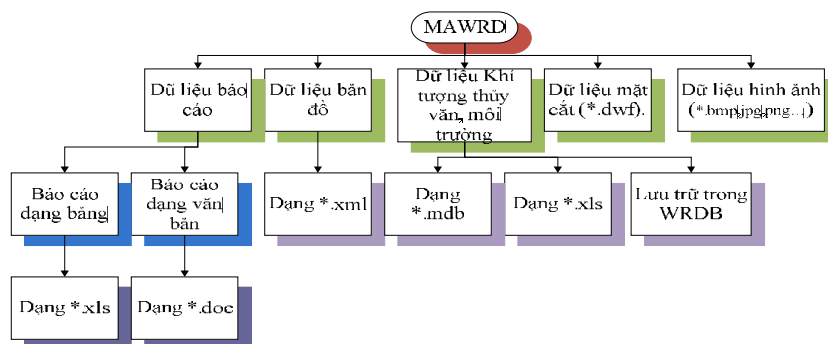
Đề tài thực hiện đã khắc phục được những khó khăn và hạn chế của việc lưu trữ và quản lý dữ liệu tài nguyên nước hiện nay bằng việc kết nối và tận dụng phần mềm WRDB 5.0 để tiến hành xây dựng dữ liệu chuẩn hóa và quản lý dữ liệu một cách hệ thống; từ đó tiến hành **lập trình** xây dựng phần mềm được đặt tên là MAWRD.

MAWRD là phần mềm quản lý, phân tích cơ sở dữ liệu cho lĩnh vực “Tài Nguyên Nước Việt Nam”, đây là phần mềm được xây dựng nhằm mục đích tạo tính kết nối với phần mềm “Water Resources Database 5.0”, liên kết với dữ liệu đã được chuẩn hóa, quản lý dữ liệu dưới dạng số và dạng bản đồ, tạo ra một số công cụ tính toán với dữ liệu.

MAWRD có giao diện thân thiện để sử dụng có nhiều tính năng hỗ trợ vào dữ liệu như tự động nhận dạng loại dữ liệu, mã các trạm KTTV của cả nước, các chức năng bản đồ và các chức năng tính toán giúp người sử dụng thực hiện nhiều phép phân tích thống kê trong thủy văn. Phần mềm này xây dựng bằng ngôn ngữ lập trình C#; đây là ngôn ngữ lập trình thuộc hệ .NET, một ngôn ngữ hỗ trợ lập trình hướng đối tượng được Microsoft phát triển. Phần mềm đã được đóng gói để dễ dàng cài đặt trên hệ điều hành Windows XP, Windows 2000, Windows Vista. Sơ đồ mô tả các chức năng của MAWRD và các định dạng dữ liệu mà MAWRD có thể quản lý được trình bày tóm tắt trong Hình 2 và 3 dưới đây:



Hình 2: Sơ đồ chức năng MAWRD.



Hình 3. Các định dạng dữ liệu mà MAWRD có thể quản lý.

Như đã nêu dữ liệu tài nguyên nước Việt Nam thường được lưu một cách không tập trung, ở nhiều định dạng khác nhau, vì vậy để tiến hành quản lý chúng một cách thống nhất, trước tiên cần chuẩn hóa chúng, vì vậy đề tài đã sử dụng WRDB để chuẩn hóa tất cả các dữ liệu thu thập được từ nhiều nguồn, nhiều định danh khác nhau về một cơ sở dữ liệu chuẩn.

**Sử dụng WRDB chuẩn hóa cơ sở dữ liệu.**

Việc chuẩn hóa cơ sở dữ liệu đã được đề tài thực hiện bằng việc xây dựng Bảng Support trong WRDB, đây là một thành phần rất quan trọng của phần mềm và là thành phần hỗ trợ cho người dùng trong tất cả các quá trình làm việc với dữ liệu. Đây là bảng quan trọng nhất. Nó chứa đựng tất cả các thông tin về hệ thống mạng lưới sông ngòi và mạng lưới trạm khí tượng thủy văn, môi trường. Bên cạnh đó, nó liên kết với tất cả các bảng dữ liệu khác trong phần mềm.

Đề tài đã xây dựng bảng Support cho phép người dùng vào, cập nhật và triết xuất thông tin một cách dễ dàng, tự động nhận dạng và kiểm tra dạng số liệu, mã trạm, với các mã trạm KTTV toàn quốc đã được WMO quy định.

Station ID	Station Name	Type	Data Freq	Latitude	Longitude	Elevation	HUC	Branch ID	River Mile	Reach ID
68739	Mỹ Hòa	KTTV								
6901	Mãng Cảnh	KTTV								
6976	Đài Lập	KTTV								
69138	Đài Lầy	KTTV								
69141	Sa Thủy	KTTV								
69702	Kom Tum(TV)	TV								
69703	Đài Mết	KTTV								
69704	Kom Lông	KTTV								
69716	Bản Đôn	KTTV								
69717	Cầu 14	KTTV								
69718	Giang Sơn	KTTV								
69719	Đức Xuyên	KTTV								

Hình 4: Bảng Support

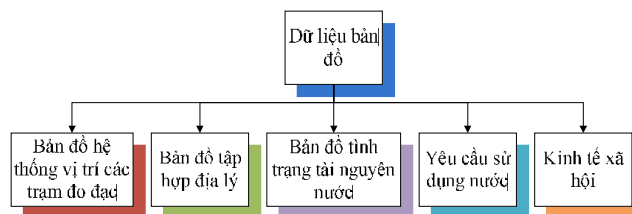
Dữ liệu sau khi đã được chuẩn hóa, sẽ được lưu dưới dạng File \*.mdb truy cập qua Microsoft Office Access. Khi người dùng thay đổi dữ liệu trong phần mềm WRDB thì dữ liệu trong file Microsoft Office Access cũng thay đổi tương ứng và phần mềm MAWRD sử dụng dữ liệu trực tiếp từ file Microsoft Office Access, vì vậy bộ ba dữ liệu này sơ bộ có tính kết nối với nhau. Tuy nhiên các tính năng của MAWRD sẽ không trùng lặp với các tính năng của phần mềm WRDB, mà hai phần mềm này sẽ có tính bổ sung hỗ trợ cho nhau, tạo nên một bộ sản phẩm quản lý cơ sở dữ liệu linh hoạt và tiện dụng cho người sử dụng.

#### Phần mềm quản lý cơ sở dữ liệu Tài nguyên nước Việt Nam - MAWRD

Phần mềm được lập này có khả năng quản lý dữ liệu về bản đồ, dữ liệu khí tượng - thủy văn - môi trường, các văn bản báo cáo, bảng biểu, dữ liệu mặt cắt, bản đồ và hình ảnh.

Với cơ sở dữ liệu khí tượng - thủy văn - môi trường, người dùng có thể sử dụng các công cụ tính toán sẵn có trong phần mềm như phần mềm WRDB 5.0, chức năng hiển thị kết quả, chức năng tính toán phân tích kết quả và xuất kết quả ra các định dạng quen thuộc.

Dữ liệu bản đồ trong phần mềm được quản lý và lưu trữ theo 5 nhóm :

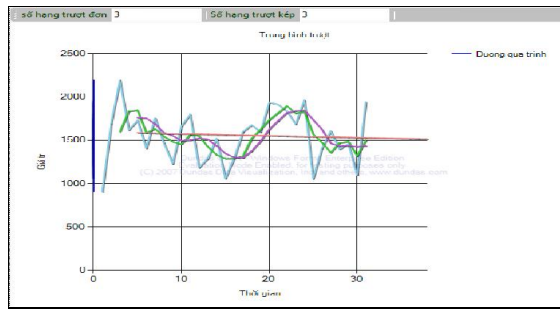


Hình 5: Sơ đồ dữ liệu bản đồ được quản lý trong MAWRD

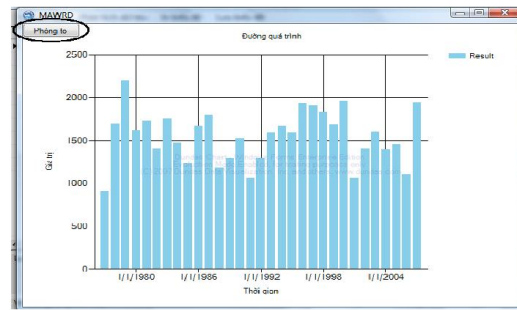
Phần mềm quản lý này có những chức năng cơ bản của một phần mềm GIS, bao gồm các chức năng phóng to thu nhỏ, kéo thả bản đồ, và hiển thị các thông tin đến người sử dụng. Ngoài ra bản đồ được xây dựng cũng có tính linh hoạt, như có thể tìm kiếm vị trí các tỉnh, các trạm đo, hay lưu vực trên bản đồ theo địa lý và theo tên của các đối tượng hiện có trên bản đồ, chức năng được thực hiện một cách nhanh chóng và tiện dụng, và đây là chức năng thể hiện tính kết nối giữa hệ thống các trạm đo đạc, về số liệu và vị trí địa lý giữa các trạm. Xét về khía cạnh khác, bản đồ có thể chia làm hai loại

chính, là các bản đồ liên kết với dữ liệu số ( hiển thị các trạm trên bản đồ, *nhóm 1* ) và các bản đồ chuyên đề ( hiển thị hình ảnh chuyên đề, các *nhóm từ 2-5* ).

Ngoài chức năng hiển thị dữ liệu, trong phần mềm này, người dùng có thể sử dụng các chức năng phân tích tính toán đơn giản được ứng dụng nhiều trong lĩnh vực tài nguyên nước như các chức năng vẽ đường trung bình trượt đánh giá số liệu về mặt chu kỳ, vẽ đường tần suất để tính toán, vẽ đường quá trình để hiển thị và phân tích số liệu và tính toán với môđun mưa dòng chảy dựa trên lý thuyết cơ bản của mô hình “NAM” được minh họa trong các hình dưới đây:



Hình 6: Biểu diễn trung bình trượt



Hình 7: Vẽ đường quá trình mưa



Hình 8: Vẽ đường tần suất.

Hình 9: Form nhập dữ liệu

Với dữ liệu mặt cắt ở định dạng \*.dwf, người dùng có thể sử dụng bàn tay để kéo thả hình ảnh hay có thể phóng to thu nhỏ tùy thích và thực hiện một số chức năng cơ bản như đang mở phần mềm Autocad.



Hình 10: Menu của Form hiển thị dữ liệu mặt cắt \*.dwf.

#### 4. KẾT LUẬN

Đề tài nghiên cứu đã lập trình và xây dựng được phần mềm MAWRD, là phần mềm quản lý, phân tích cơ sở dữ liệu cho lĩnh vực “Tài Nguyên Nước Việt Nam”, đây là phần mềm được xây dựng nhằm mục đích tạo tính kết nối với phần mềm “Water Resources Database 5.0”, liên kết dữ liệu đã được chuẩn hóa, quản lý dữ liệu dưới dạng số và dạng bản đồ, kết nối với một số công cụ tính toán cơ bản với dữ liệu hiện có. Giao diện thân thiện dễ sử dụng có nhiều tính năng tính toán giúp đỡ cho người sử dụng. Đặc biệt hơn phần mềm đã có sẵn một nền chuẩn để nhập số liệu bao gồm mã của tất cả các trạm KTTV do Tổ chức Khí tượng Thế giới quy định cho Việt Nam. Với nền chuẩn này, phần mềm có thể dùng để quản lý dữ liệu cho tất cả các lưu vực, các tỉnh, vùng khác nhau của Việt Nam để

xây dựng và quản lý số liệu ở mọi định dạng, biểu mẫu khác nhau. Phần mềm đã được đóng gói và dễ dàng cài đặt trong môi trường Windows.

## 5 TÀI LIỆU THAM KHẢO CHÍNH

1. Hoàng Thanh Tùng, Ngô Lê An, Trịnh Quang Toàn (5-2009). Chuyên đề Xây dựng Chương trình Quản lý CSDL phục vụ Khai thác Nguồn nước Tỉnh Ninh Thuận, Bình Thuận. Viện Tài nguyên nước và Môi trường Đông Nam Á.
2. Hội đồng Tài nguyên nước Quốc gia (2004). *Atlas Tài nguyên nước Việt Nam*.
3. Giới thiệu về phần mềm WRDB tại Website: <http://www.wrdb.com/wrdb4.5/index.htm>

### Summary

*The research “Standardizing database and developing software for analysis and management of water resources database of Vietnam” deals with outstanding difficulties and constrains in the management of water resources database of Vietnam. The outstanding difficulties and constrains includes: i) data has being stored un-synchronously with different format, ii) storing type is not unified, iii) it is difficult to store data in standard type and to manage data in a system way, iv) it has not any clear assessment on the quality of data, v) it is not possible to connect with observation system and with their locations. For the mentioned purposes, this research has developed a system of several modules including the data-standardizing module, software for management and analysis of water resources database of Vietnam with functions dealing with the above mentioned difficulties and constrains as well as supporting to the analysis, estimation of hydrology and connecting to the common hydrological and hydraulic models.*