UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE Mathématiques, Sciences de l'Information et de l'Ingénieur

[Uds – INSA - ENGEES]

THÈSE présentée par :

Gula TANG

soutenue le **30 Mai 2016**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Electronique / Télédétection

Research on Distributed Warning System of Water Quality in Mudan River Based on EFDC and GIS

THÈSE dirigée par :

M. Zhao-Liang Li	Directeur de Recherches CNRS, France
M. Jiulin Sun	Directeur de Recherches UCAS, Chine

RAPPORTEURS :

M. José Sobrino	Professeur, University de Valencia, Espagne
M. Zhihao Qin	Directeur de Recherches CAAS, Chine

AUTRES MEMBRES DU JURY :

Mme Françoise Nerry	Directrice de Recherches CNRS, France
M. Guangjian Yan	Professeur de l'Université Normale de Beijing, Chine

Acknowledgement

As a joint student, my research has been done at University of Strasbourg in France and University of Chinese Academy of Sciences in China. Throughout these years, I am always very grateful to all the guidance, helps and encouragements that I have received.

First and foremost, I owe the deepest gratitude to my Ph.D. supervisor, Dr. Zhao-Liang Li (Director, CNRS and University of Strasbourg) and Prof. Jiulin Sun (Academician, the Chinese Academy of Science, University of Chinese Academy of Sciences) for giving this opportunity to do my Ph.D. research and for their enthusiasm, inspirations and valuable advices in every important step of my work. I have acquired so much knowledge and obtained so many good ideas from the discussions with them; this thesis would not be completed without their guidance and help.

I am grateful to Dr. Françoise Nerry (Director, CNRS and University of Strasbourg) for her support and help.

I would like to thank my dear friends: Dr. Alexis Ren, Dr. Lei Lin, Dr. Sibo Duan, Dr. Nicolas Fan, Dr. Yuan Liu, Dr. Hongyuan Huo, Dr. Sébastian Leng, Dr. Zhuoya Ni for their support and concern. Simultaneously, I also thank Dr. Simon Yuan (Ecole Normale Superieure), Dr. Michelle Song and Dr. Léa Yang (Université Claude Bernard Lyon 1) for their encouragement and help. Especially, I also express my gratitude to Dr. Jun Chen, Dr. Guozheng Wu, Ms. Yi Wen, Dr. Georges Shi (Université Paris Diderot) for their endorsement and assistance during my writing the dissertation.

I am deeply grateful to my teachers, Ms. Yaping Yang, Mr. Huazhong Zhu, Mr. Juanle Wang, Mr. Yunqiang Zhu (Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences) for their concern, patience, optimism and help. And their excellent qualities will always inspire me.

I take this opportunity to express my sincerest love and thanks to my family and for the love and support they give me, and for all the sacrifices they have done for me. Finally, I owe the deepest gratitude to my wife, Ping Wen (Université Paris Diderot). I can't finish the dissertation without her inspiration, encouragement and endorsement.

My research was financed by Major Science and Technology Program for Water Pollution Control and Treatment in China, Mudan River Water Quality Integrated Security Technology and Engineering Demonstration Project (No.2012ZX07201002), China Scholarship Council, Key Program of the fundamental research of the science and technology of China (No. 2013FY110900), and the foundation condition platform of national science and technology of China (No. 2005DKA32300).

Contents

List of Figures

List of Tables

Résumé

Les événements de pollution de l'eau qui se déroulent en Chine ont donné l'alarme d'une situation désastreuse, les incidents de pollution causant de larges dégâts et des pertes économiques énormes. Il est donc vraiment nécessaire de réaliser des études sur le Système de simulation et d'avis précoce d'alerte de la qualité de l'eau dans le but de renforcer la capacité de prévenir la pollution de l'eau, de faire diminuer les dommage écologiques pour l'eau, les animaux et les humains, de réduire les pertes économiques et d'améliorer la qualité de vie. Ainsi, par le système informatique précité, la variation de la répartition spatio-temporelle de la qualité de l'eau peut être précisément simulée, prévue, évaluée afin que le gouvernement puisse appliquer des politiques raisonnables et adopter des mesures adéquates dans le domaine de la gestion de l'eau.

La rivière Mudan est une rivière importante dans les régions froides du Nord-Est de la Chine et est un affluent de rivière Songhua après réunion avec rivière Heilong, qui se jette finalement dans la rivière de l'Amour en Russie. Donc la qualité de l'eau dans la rivière Mudan est une préoccupation non seulement au niveau local et régional, mais aussi au niveau international. Chaque année en hiver, la rivière se recouvre de glace pendant presque cinq mois. C'est pourquoi la création et le fonctionnement du Système de simulation et d'avis précoce d'alerte de qualité de l'eau de la rivière Mudan font face à des nombreuses difficultés. Du fait de la situation particulière de cette rivière, il est urgent d'intégrer la technique des systèmes d'information géographique (SIG) dans les modèles de simulation de la qualité de l'eau, pour bien surveiller des sorties d'égouts au long de la rivière notamment dans les cas de dégorgement surchargé et décharge non déclarée, pour diminuer les possibilités d'incidents de pollution, pour réduire les pertes économiques, pour finalement soutenir la gestion quotidienne sur l'environnement et la qualité de l'eau de la vallée correspondante, et aussi pour aider à établir des décisions administratives en fournissant les informations importantes en cas d'urgence. Ainsi, on pense que cette recherche sur le Système de simulation et d'avis précoce d'alerte

de la qualité de l'eau pour la rivière Mudan aura des impacts particuliers, par exemple pour renforcer les études sur les techniques de la protection des eaux dans la région froide du Nord de la Chine, et dans le but d’aider à éviter les discordes internationales.

Nous voulons donc créer un système pratique de simulation et d’avis précoce d’alerte pour que la distribution spatio-temporelle de la qualité de l'eau durant les périodes de couverture glaciaire et d'eaux libres soit simulée précisément en utilisant des données depuis un petit nombre de sites d'observation pour permettre de comprendre la variation spatiale de polluants sur le parcours de la rivière et leur conséquences. Selon les résultats de la simulation, non seulement la qualité de l'eau dans la rivière pourra être évaluée et visualisée par des tableaux, des graphiques et des cartes visuelles, mais l’avis précoce d’alerte pourra être lancé au cas où les indices de pollution dépasseraient des critères prédéfinis ou si des événements de pollution survenaient soudainement.

Pour créer un tel système, nous avons rencontré plusieurs défis.

La première difficulté concernait le fait que les données d'observation disponibles pour la qualité de l’eau de rivière étaient très limitées. Il a été difficile d’étalonner les paramètres requis pour le système de modélisation du fait de ces données limitées. De plus, pour que la répartition spatio-temporelle de la qualité de l’eau de toute la rivière puisse être évaluée et que l’influence des eaux usées depuis des sorties d'égouts puisse être surveillée, les données disponibles de la qualité de l’eau depuis un petit nombre de sites d'observation n’étaient pas suffisantes pour effectuer une simulation scientifique complète. Afin de contourner cet inconvénient, on a créé le système de simulation et d’avis précoce d’alerte sur la base du mécanisme de modèle EFDC (The Environmental Fluid Dynamics Code) qui avait été recommandé par US EPA (l'United States Environmental Protection Agency) pour simuler les processus chimiques et physiques des polluants qui diffusaient et étaient transportés sur le cours de la rivière.

La deuxième difficulté concernait le problème de calcul du système de modélisation, comme un système de modélisation distribuée, contenant des milliers de grilles à calculer concurremment pour les scénarios simulés. Le travail de calcul du système était une lourde tâche qui devait durer longtemps et qui utilisait une part énorme des ressources de calcul des serveurs. Donc il était très important de trouver une solution pour améliorer l’efficacité de l’analyse des scénarios et pour visualiser en temps réel la propagation spatio-temporelle de la pollution de l’eau. Si cela fonctionne, le temps de la réaction de la simulation serait considérablement réduit en cas d’un traitement d'urgence. Pour résoudre ce problème, on a introduit le SIG (*système d'information géographique*) pour visualiser la qualité de l’eau, et on a réalisé un calcul distribué de modèle, avec une efficacité très élevée dans différentes analyses de scénarios de changement de qualité de l’eau dans la rivière.

La troisième difficulté concernait la prévision précise du changement spatio-temporel de qualité de l’eau pour que la meilleure action puisse être prise le plus tôt possible pour traiter les incidents de pollution potentiels et pour diminuer les pertes quand la pollution se sera vraiment produite. Dans cette recherche, divers scénarios sont développés pour offrir une bonne estimation de la dynamique de qualité de l’eau dans la rivière.

La thèse est structurée en 7 chapitres. Dans le chapitre I, on décrit le contexte et la signification de cette étude, on examine les progrès actuels sur la surveillance de la qualité de l'eau dans les régions froides, aussi que les situations présentes des recherches internationales sur le Système de simulation et d’avis précoce d’alerte de la qualité de l'eau. On a procédé aux études comparatives sur la force hydrodynamique et la qualité de l'eau de la même rivière durant les périodes de couverture glaciaire et d'eaux libres. Nous avons trouvé que le modèle 2D de qualité de l’eau devait être encore développé en le comparant avec ce que les chercheurs européens ou américaines avaient déjà réalisé. Beaucoup des questions méritent d'être posées, comme par exemple, les différences sur le processus hydrodynamique et le processus de déplacement et transformation des polluants dans la période de couverture glaciaire et la période d'eaux libres et les raisons qui ont causé ces différences. En conséquence, l'étude de la simulation de la qualité de l’eau dans la région froide du nord, qui prend la rivière Mudan comme exemple, notamment la qualité de l’eau dans la période de couverture glaciaire a une grande importance pour bien construire de système de simulation et d’avis précoce d’alerte qui pourra être utilisé dans une région froide similaire.

Dans ce chapitre I, on présente aussi la méthodologie et le cadre des recherches. Une revue de la documentation pertinente pour comprendre les progrès actuels sur la modélisation de la qualité de l’eau, les nouveaux développements des modèles et les particularités des différents modèles pour le processus de la modélisation est réalisée. Les données requises, en particulier les données de surveillance de la qualité de l'eau, ont été prélevées pour faire cette analyse. Dans ce chapitre, on analyse ensuite la méthode et les techniques pour l’avis précoce d’alerte, on compare les avantages et les inconvénients des différentes méthodes et les caractéristiques des modèles existants, afin de choisir le meilleur modèle comme l’outil de base pour établir le système de modélisation pour la rivière Mudan. Aussi on analyse les sites de surveillance de la qualité de l'eau de la rivière Mudan avant la construction des modèles. Le modèle hydrodynamique a été construit, l’étalonnage et la vérification des paramètres du modèle hydrodynamique ont été réalisés. Le même processus pour le modèle de qualité de l’eau est également réalisé. Ce dernier modèle qui est déjà vérifié pour la rivière Mudan est très important pour la construction de système distribué de simulation et d’avis précoce d’alerte. A ce moment là on a pris en considération l’intégration des techniques de SIG avec le modèle de qualité de l’eau de la rivière Mudan, la conception du cadre distribué, le système des fichiers répartis, le calcul distribué et la méthode d’ordonnancement équilibré, ensuite le problème de la conception de cadre de système de l’avis précoce de alerte de qualité de l’eau et la mise en œuvre des technologies décisives pertinentes. Le préavis d’alerte habituel de la qualité de l’eau et la prévision d’incident de pollution soudaine utilisent le modèle de qualité de l’eau de la rivière Mudan qui était déjà vérifié. Dans cette étude, Il y a aussi des technologies décisives développés dans le système. En dernier lieu on effectue la simulation des scénarios de qualité de l'eau et des incidents de pollution soudaine pour démontrer que les modèles et le système sont bien construits.

Dans le chapitre II, on analyse les phases du développement, la définition et la classification des modèles de qualité de l'eau et on discute les moyens techniques usuels de la simulation et du préavis d'alerte de qualité de l'eau, y compris les modèles mécanistes et non-mécanistes. Le modèle non-mécaniste comprend principalement la méthode statistique mathématique, la théorie des systèmes gris, la méthode de réseau neuronal etc. le modèles mécaniste comprend les outils de la modélisation, par exemple CE-QUAL-ICM, WASP, EFDC, MIKE, DELFT3D, QUAL2k etc. Dans ce chapitre, on compare les avantages et inconvénients des méthodes de préavis d'alerte de qualité de l'eau fondés sur les modèles mécanistes et sur les modèles non-mécanistes, en tenant compte de l'objectif de la recherche. On choisit finalement le logiciel EFDC comme base pour créer le modèle de qualité de l'eau de la rivière Mudan. Ce logiciel a les facultés suivantes : 1) Comme outil de la simulation de l'eau de surface, EFDC est recommandé fortement par le EPA des Etats-Unis. Sa précision sur la simulation hydrodynamique et la qualité de l'eau obtient des reconnaissances considérables en milieu académique. 2) Le moyen de saisir les fichiers est facile à comprendre et le couplage avec des modules d'hydrodynamique, de la qualité de l'eau et de sédiments est aisé. Les transformations entre les ports de différents modèles peuvent être négligés. 3) Comme un logiciel à code source ouvert, EFDC n'est pas seulement pratique pour intégrer des modèles et des composants fonctionnels de SIG ainsi que leur développement secondaire, mais il est favorable pour créer un système distribué de simulation et préavis d'alerte. Dans le but de créer un tel système, on étudie les éléments suivants : la simplification de calcul intensif du modèle, l'amélioration de l'extensibilité de la simulation du modèle, la réduction de dépense de hardware, le calcul dynamique élastique de modèle de qualité de l'eau. A la fin de ce chapitre, on présente certains des concepts de base, l'équation et la solution numérique de cet outil de modélisation.

La construction de modèle de qualité de l’eau de la rivière Mudan est la partie de base du procédé de l’établissement de système de simulation et l’avis précoce d’alerte durant lequel on doit prendre en considération la situation particulière de la région que l’on étudie. Dans le chapitre III, on analyse la situation courante de qualité de l’eau de la rivière Mudan et les données observées de 2000 à 2014 ont été utilisées. Cela montre que, excepté pour l’année 2008 où la qualité de l’eau était plutôt de catégorie III, la qualité de l’eau était principalement de catégorie IV. De 2009 à 2013, la proportion de l’eau de catégorie III a augmenté, d’un peu plus d'un tiers, la qualité de l’eau s'étant améliorée. De 2013 à 2014, la qualité de l’eau s’est dégradée, surtout en 2013, la proportion de l’eau de catégorie IV est montée jusqu’à 85.71%, Elle atteignait alors le pic des quinze dernières années. En 2014, la proportion de l’eau de catégorie V arrivait à 16.67%, et une partie de l’eau même appartenaient la catégorie inférieure à la catégorie V.

Selon les données de topographie et des données de surveillance de la qualité de l'eau, on précise dans ce chapitre comment établir le système de réseau fluvial et instaurer les conditions limites. Enfin un nouveau 2D modèle de hydrodynamique et de la qualité de l'eau est établi, la simulation de la concentration de COD et NH_3N durant la période de couverture glaciaire et la période d'eaux libres de la rivière Mudan rivière est réalisée. La *grille curviligne orthogonale est utilisée pour diviser le plan d’eau en 4207* unités, la matrice est de 864 lignes x 5 colonnes, la dimension de chaque unité mesure entre 24×43.6m et 176.9×241.3m. La hauteur du lit du fleuve diminue depuis 245.7m dans la section Xige jusqu’à 215.1m dans la section Pont de Chaihe. La condition limite de débit de modèle comprend le débit d’eau venant de l’amont dans la section Xige, la quantité de la décharge de l’eau vers l’aval dans la section Pont de Chaihe, les flux coulent depuis le tributaire de la rivière Hailang dans la rivière principale, et le volume de l’évacuation de l'eau usée depuis des sorties d'égouts le long de la rivière. Pour le segment de flux concernant la simulation, seulement une station de surveillance sur l'hydrologie nommée S2 est établie dans la section de la rivière Hailang, les stations de surveillance sur la qualité de l’eau étant établies dans la section Xige et la section Pont de Chaihe, donc le débit d'eau venant de l'amont dans les sections de rivière Hailang et Xige est calculé depuis les données de débit observé recueillies dans la station S2. Selon les proportions prises par les zones de drainage contrôlées à l'amont de la section Xige et de sous-flux de la rivière Hailang, respectivement, les données de débit recueillies dans la station S2 sont attribuées à la section Xige et à l'entrée de la rivière Hailang. Parmi eux, les flux de la section Xige occupe les 2/3 de celle de la station S2, les flux de la rivière Hailang rejetées dans la rivière Mudan compte pour 1/3 de la même station S2. En ce qui concerne la condition limite du Pont de Chaihe, la condition limite ouverte est adoptée, à savoir la condition concernant le niveau d'eau. Les données de niveau d'eau peuvent être acquis par le moyen suivant : soustraire la distance (entre la hauteur du lit du fleuve de la section de station S2 et celui de la section Pont de Chaihe) du niveau actuel de l'eau de la station S2.

Il convient de noter que la décharge des eaux usées depuis des sorties d'égouts de la rivière est une donnée de débit observé recueillie actuellement. La condition limite de concentration comprend la concentration des polluants dans l'eau qui viennent de l'amont de la section Xige, la concentration des polluants dans l'eau qui viennent de la rivière Hailang, et la concentration des polluants depuis des sorties d'égouts le long de la rivière. Pendant le processus de vérification du modèle, quatre sections sont choisies, y compris le Pont de Wenchun, Hailang, le Pont de Jiangbin, et le Pont de Chaihe. Habituellement, les surveillances de qualité de l'eau s'effectuent pour chaque section au début des mois suivants: janvier, février, mai, juin, juillet, août, septembre et octobre. Parmi eux, les données de janvier et février représentent la période de couverture glaciaire et les mois restant représentent la période d'eaux libres. Dans la région de ce que concerne la modélisation, onze sorties d'égouts sont pris en considération. D'après les données mesurées, il n'y a pas de données successives de surveillance des flux depuis ces onze sorties d'égouts, il existe seulement les données concernant la totalité de décharge des eaux usées. Par conséquent, une valeur constante est adoptée pour décrire le débit des eaux de chaque sortie d'égouts. Étant donné que la concentration de polluants est surveillée une fois par trimestre, les valeurs observées sont utilisées pour la condition limite de la concentration. Chaque année, environ 60 millions m^3 d'eaux usées depuis ces 11 sorties d'égouts sont évacués dans la rivière Mudan, la concentration de polluants de quelques sorties est élevée et même dépasse le standard national d'évacuation des eaux usées polluant ainsi la rivière.

De plus, nous avons étalonné et vérifié les paramètres impliqués, y compris le coefficient de dispersion horizontale, le coefficient de diffusion turbulente verticale, le taux d'atténuation de COD et NH_3N, la *rugosité du lit etc. Selon notre recherche et notre analyse,* des conclusions peuvent être tirées de la manière suivante: 1) les résultats des recherches montrent que parmi les quatre sections, la différence entre la valeur analogique et la donnée mesurée de la concentration de COD *est la plus grande dans la section* Pont de Chaihe, l'erreur relative étant de 18.43% et la plus petite différence concerne *la section* Pont de Wenchun avec une erreur relative de 5.86%. Quant à la concentration de NH_3N pour les quatre sections, la plus grande erreur relative est de 39.58% et la plus petite de 14.88%. 2) le taux d'atténuation de COD et NH_3N pendant la période de couverture glaciaire est plus bas que pour la période d'eaux libres. Selon les résultats de nos recherches, dans la rivière Mudan, le taux d'atténuation de COD et NH_3N durant la période d'eaux libres est de 0.03 par jour et de 0.05 par jour, et pendant la période de couverture glaciaire, de 0.01 et 0.02 par jour. 3) Dans cette recherche, la *rugosité adoptée pour* la période de couverture glaciaire est de 0.043 et 0.035 *pour* la période d'eaux libres. Les résultats de la simulation nous montrent que ces deux paramètres ont été choisis de manière appropriée. 4) D'après le taux d'atténuation de COD et NH_3N des autres rivières en Chine, celui de la rivière Mudan est relativement moins élevé. Cela peut être dû à la faible température annuelle moyenne de cette rivière qui est située dans une région froide. Bien que les valeurs simulées de COD et NH_3N montrent une excellente cohérence avec leurs données mesurées, certaines erreurs existent encore. Les principales raisons qui causent ces erreurs de calcul de la concentration de polluant sont suivantes : 1) L'erreur due au fait de généraliser la hauteur de lit de rivière au même niveau; 2）La variabilité des sources de pollution qui se déversent dans le cours d'eau depuis des réseau urbains de drainage des eaux pluviales et des canalisations d'eaux usées. 3) Les pollutions diffuses agricoles qui se dispersent de chaque côté du courant de la rivière et qui présentent de réelles difficultés pour bien les mesurer.

Après l'analyse et la vérification du modèle, dans son ensemble, il est possible de le coupler avec un SIG et finalement construire le Système de simulation et d'avis précoce d'alerte.

Dans le chapitre IV, on analyse la méthode d'intégration du modèle de qualité de l'eau de la rivière Mudan dans le système de préavis d'alerte, avec un SIG , ainsi que l'architecture distribuée du calcul dynamique élastique de modèle de qualité de l'eau. Selon les caractéristiques de calcul du modèle de mécanisme de qualité de l'eau, une méthode d'intégration étroitement couplée est adoptée dans cette recherche. En tenant compte de la simulation spatio-temporelle de la qualité de l'eau et le bon fonctionnement exigé lorsque la pollution de l'eau advient, on étudie une méthode de calcul distribué, pour distribuer les calculs du modèle dans le cluster de serveurs avec des nœuds élastiques, on précise l'architecture distribuée de modèle de qualité de l'eau avec le calcul dynamique élastique. Dans une telle architecture, on étudie le procédé de communication entre les nœuds du serveur, le mode de communication entre les serveurs et les clients. En outre, étant donné la transmission de fichiers de projet entre les nœuds élastiques de serveur ou la transmission avec le côté client, en combinant avec le système de fichiers distribué existant, on étudie les méthodes de gestion de stockage binaire de fichiers de projets de simulation de qualité de l'eau, on détermine le partage distribué de programme de simulation ainsi que le procédé de mémorisation des résultats de la simulation du modèle. Cette approche peut nous aider à simplifier considérablement la transmission des données entre les nœuds de serveur, afin d'améliorer l'efficacité du calcul distribué de modèle sous la même condition de réseau. Les nœuds élastiques de serveurs distribués et le partage de fichiers de manière distribuée sont les bases importantes pour le calcul distribué de modèle de qualité de l'eau de la rivière Mudan.

Après l'introduction de la programmation de modèle distribué, on discute dans le chapitre V l'équilibrage de la charge de la tâche planifiée de calcul distribué, les principaux indices d'évaluation des nœuds élastiques, y compris le stockage disponible des nœuds, l'utilisation de la CPU, la charge de transmission du réseau, le stockage de fichiers et la mémoire disponible etc. Ensuite, dans ce chapitre on discute le calcul avec l'équilibrage des charges de stockage des données, l'ordonnancement de l'équilibrage de charge dynamique pour le calcul avec des nœuds élastiques, on explique aussi l'ordonnancement de l'équilibrage de charge de Map/Reduce pour le téléchargement du programme de simulation. En outre, on analyse le calcul distribué de la simulation de qualité de l'eau, y compris le calcul parallèle des systèmes multi-modèles d'un seul nœud, la stratégie de l'ordonnancement de l'équilibrage de charge de Map/Reduce. Le rôle du *pool threads* du calcul de nœuds élastique est d'améliorer la capacité des simulations de multi-projets pour un seul nœud. Dans l'ordonnancement de l'équilibrage de charge de Map/Reduce, l'opération de Map concerne le processus de la simulation des nœuds élastiques, l'opération de Reduce concerne les procédés du téléchargement et de l'affichage des résultats de la simulation depuis un client. Pour assurer la fiabilité des nœuds élastiques finalement retenues, quand le serveur maître reçoit une demande de calcul de la simulation de qualité de l'eau depuis un client, il choisit quelques serveurs d'après les informations disponibles et les envoie au client. Donc le client pourra obtenir les dernières informations sur le nombre de tâches sur les nœuds. Sur cette base, les nœuds sont classés par ordre croissant, le premier étant généralement l'optimum. Afin d'améliorer la fiabilité de l'exécution des tâches, au moment de choisir les meilleurs nœuds, le cycle de la vérification est adopté et les informations d'échec depuis des nœuds qui exécutaient des tâches sont traitées.

Dans le chapitre VI, on étudie la structure logique et le cadre de fonctionnement du système de préavis d'alerte de qualité de l'eau. Le système est constitué logiquement par la couche physique, la couche de données, la couche application et la couche de présentation, que l'on les expose respectivement. Ensuite, les approches pour réaliser certaines techniques clés en rapport avec le système de préavis d'alerte, y compris la base de données du système, l'évaluation de la qualité de l'eau, l’analyse sur la tendance de la qualité de l'eau, le préavis d’alerte du dépassement des critères autorisés, la prévision de routine de la qualité de l’eau, l’alerte précoce à la pollution soudaine sont discutées. Les deux dernières techniques s’appliquent aux fonctions centrales de système qui est fondé sur la base du modèle hydrodynamique et de modèle de qualité de l’eau. Enfin le prototype de système de préavis d’alerte de qualité de l’eau, l’outil de gestion des nœuds élastiques, l’outil de gestion du serveur maître sont présentés dans ce chapitre.

On travaille sur des scénarios d'application dans le cas d’une situation ordinaire et dans le cas de la pollution soudaine et on traite la méthode de mise en œuvre du système distribué. Par l’outil <créer un nouveau projet > du système, on établi des projets de simulation des scénarios, quand les projets commencent à fonctionner, les opérations en arrière-plan sont consécutivement les suivantes: 1) Prendre le nombre de projets de simulation en cours d'exécution pour chaque nœud d’exécution comme une base d’estimation, trouver le nœud optimal qui a le nombre le moins élevé d'exécution de projets. 2) Renseignez les fichiers locaux des projets par projets ID. 3) Lire les données du projets locaux et les transmettre au nœud d'exécution optimal. 4) Le nœud d’exécution reçoit des projets, les enregistre dans le catalogue correspondant et procède à une identification unique en nommant des fichiers avec les projets ID. 5) Ce nœud d’exécution fait une copie de fichier sur le modèle de la qualité de l’eau de la rivière Mudan à partir de la matrice, et procède à une identification unique en nommant de fichier avec les projets ID. 6) Ce nœud d’exécution lit des données de projet et les substitue au fichier déjà copié sur le modèle de la qualité de l’eau de la Mudan. 7) Ce nœud d’exécution essaie d’opérer le modèle déjà substitué de qualité de l’eau de la rivière Mudan avec une forme de *multithreading*. 8) Si l’opération est réussie, le chemin où le projet est situé sera enregistré dans le système de fichiers distribués. 9) Si parmi les étapes précédentes de 3 à 7, l’information concernant l’erreur est renvoyée, le système permettra d'éliminer ce nœud et de retourner à l’étape 1 jusqu'à ce que le système fonctionne avec succès. 10) Après que l’opération de modèle de prévision de qualité de l’eau soit réussie, ce nœud d’exécution enregistrera le fichier *log* terminé après le fonctionnement du modèle dans le système de fichiers distribué. 11) Télécharger les résultats de simulation. Par le système de fichiers distribué, on interroge de quel nœud d'exécution et par quel chemin les données de simulation de ce projet sont situées, puis on effectue le téléchargement correspondant. 12）Après que les fichiers sur les résultats de la simulation soient téléchargés, on les réceptionne dans le SIG côté client, et on affiche la dynamique spatio-temporelle.

On utilise le serveur déployé Ali de cluster de serveurs afin de diminuer la complicité du test. La configuration la plus basique est adoptée pour tous les serveurs avec des nœuds élastiques. Quand on utilise le mode nœud-unique pour la simulation et la prévision, chaque projet monopolise les serveurs, le temps total est la somme des consommations de temps de ces serveurs, à savoir 4.48h. Quand le mode de la calcul de modèle distribué est adopté, le temps total dépend du projet de simulation de scénario qui dure le plus longtemps, soit 1.1579h. Le mode du calcul distribué améliore donc considérablement l'efficacité du calcul de la simulation des scénarios. En même temps par ce mode, quand la planification de projet de simulation des scénarios change, le nombre de serveurs peut augmenter ou diminuer selon les besoins. On obtient non seulement l'efficacité de la simulation et la prévision élevée, mais on réalise des économies sur l'utilisation de l'équipement matériel des serveurs. Si la largeur du cours d'eau concernant la simulation et la prévision est plus importante et si la période est plus longue, le nombre de grilles de modèle est plus nombreux et les avantages du calcul distribué de modèle seront plus évidents.

Le chapitre VII est un résumé de la thèse et présente les conclusions. Il analyse les points innovants du mémoire et donne les perspectives de ce travail. L'étude présente trois points innovants: 1) Un système bidimensionnel de simulation et d'avis précoce d'alerte pour la qualité de l'eau de la rivière Mudan est établi, la propagation de polluants dans la rivière est différente entre la période de couverture glaciaire et la période d'eaux libres et cette différence est prise en considération dans ce modèle. Les coefficients de diffusion convective, la rugosité du lit de la rivière, et le taux d'atténuation ont été soigneusement vérifiés et calibrés pour le

modèle en utilisant les données disponibles depuis les sites d'observation limitées. La vérification et l'étalonnage garantissent l'exactitude de la simulation. Une panoplie complète de fonctions utiles en termes d'applications pratiques a été développée dans le système pour répondre aux différents objectifs de la gestion de la qualité de l'eau notamment le préavis d'alerte de dépassement des critères prédéfinis de COD et NH_3N, ou pour d'autres indices similaires de pollution, comme l'oxygène dissous, le phosphore total et l'azote total dans le cours d'eau de la rivière Mudan. 2) Une architecture de calcul efficace a été développé et structurée afin de réduire les tâches de calcul et la simulation arrive à un niveau acceptable. On réalise la mémorisation, le partage et le téléchargement distribué de projets de simulation de modèle de la qualité de l'eau. On réalise le calcul dynamique élastique de modèle et améliore l'efficacité de l'intervention d'urgence. 3) Un prototype du Système de simulation et d'avis précoce d'alerte de la qualité de l'eau est développé. Le système permet une description quantitative et l'analyse visualisée de la tendance de diffusion convective et de la variation des polluants dans le cours d'eau. En combinant l'évaluation de la qualité de l'eau, l'analyse des tendances, la simulation quotidienne et la prévision d'alerte rapide de la qualité de l'eau, on réalise l'analyse de la dynamique spatio-temporelle de qualité de l'eau et améliore considérablement la capacité du suivi et de la gestion fine de la qualité environnementale de l'eau de la rivière Mudan.

Mots clés: préavis d'alerte; modèle hydrodynamique; qualité de l'eau; simulation bidimensionnelle; calcul distribué; analyse de scénarios; SIG

中文摘要

众多水污染事件的惨痛经验告诉我们：水污染事件的后果是灾难性的，影响是深远的，代价是沉重的。为了加强水污染预测预警能力，减轻对水生态、动物和人体的伤害，降低经济损失和提升人居环境质量，需要开展水质的模拟预警技术研究，结合 GIS 技术集成水质模拟模型，模拟、预测、评价水质的变化趋势，监控超排漏排河流排污口污水排放状态，支撑流域环境水质管理日常业务，强化水环境的监控预警，尽量减少水污染事故的发生，强化污染发生前后的水质预测预警能力，为水污染应急提供决策信息支持，减少水污染发生后的损失，具有重要意义。

对处于中国北方寒冷地区的牡丹江而言，河流水质模拟预警技术是水环境安全的有力保障，但每年有近一半的时间处于冰封期，要基于软件构建一套实用的水质模拟预警系统，利用有限的河流断面实测数据，既能在非冰封期模拟预测河流的水污染时空分布，又能在冰封期监控河流水污染传播情况，达到准确掌控河流水质空间分布状况，预测水质变化趋势，预警常规监测污染指标超标和突发污染事故发生的目的，构建过程将面临诸多挑战。挑战之一是用什么方法可以仅仅利用有限的断面监测数据，就能够掌握整条河流的水质时空分布状况，监控排污口污水对河流水质造成的影响；挑战之二是当污染事故发生时，如何提高水污染时空传播模拟的效率，以缩短污染事故应急响应的时间；挑战之三是如何对河流的水质进行预测，以提前对可能造成的水污染做好准备，避免水污染事故的发生，减少水污染造成的损失。

本论文以构建中国北方寒冷地区牡丹江流域水质模拟预警系统研究为目标，针对以上三个方面的挑战内容，研究了牡丹江水质现状、水质机理模型构建技术、水质机理模型与 GIS 的集成技术、模型的分布式计算和水质模拟预警系统相关核心技术的实现等内容。牡丹江为松花江第二大支流，发源于吉林长白山的牡丹岭，干流沿岸分布有多个居民区和工业区，大量的工业废水和生活污水通过各种方式排入了牡丹江中，使得牡丹江的水质状况堪忧。研究结果显示，牡丹江干流城市段水体呈现有机污染特征，主要的超标因子为化学需氧量（COD）和氨氮（NH_3N）。牡丹江流域的水污染可能会直接汇入松花江干流，从而进一步污染出境河流黑龙江的水质。出境河流的污染可能引发中俄间的环境外交争端，造成严重的国际影响。牡丹江流域水质模拟预警系统研究的意义可以归纳为以下几点：1）加强水质预警预测能力，防止国际争端；2）

响应政府规划，加强完善水质监测预警体系，保障水质安全；3）提高水质预警预测能力，提高水质管理和应急反应的效率。

论文一共分为七章。第 1 章分析牡丹江水质分布式预测预警系统研究的背景和意义，讨论 GIS 技术支持下的水质预警系统在国际上的研究进展情况，阐述寒冷地区河流水质模拟预测技术的研究现状，描述水质模型的分布式计算，指出已有研究的当前国际水平和存在的不足。在此基础上提出本研究的主要内容、研究的技术路线和论文各个章节的安排情况。在中国对于同一河流进行冰封期和非冰封期水动力与水质过程进行对比研究，构建二维水质模型的研究相对于欧美等发达国家仍然需要加强。冰封期与非冰封期河道的水动力过程与污染物迁移转化过程有哪些差异，造成这些差异的原因是什么，主要的影响因素或指标有哪些，这些问题都非常值得研究。因此，对以牡丹江为代表的中国北方寒区河流水质，特别是冰封期水质进行模拟研究就显得特别重要，从而更好地为构建与牡丹江类似的寒区河流水质预测预警系统服务。为了实现论文的研究目标，本章还设计了研究技术路线：在相关文献调研的基础上，准备处理需要的研究数据和研究材料。然后研究水质模型预警技术概况，分析不同的水质模型的特点，选择合适的方法来构建牡丹江的水质模型。之后分两个主要的研究路线并行研究，其一是分布式预警系统这条技术路线上，需要考虑 GIS 与水质模型的集成，分布式框架设计，分布式文件系统，以及分布式计算和均衡调度方法；然后是水质预警整体系统框架的设计以及相关核心技术的实现问题。核心技术中的常规水质模拟预警和应急水质模拟预警会使用牡丹江水质模型，因此另外一条技术路线需要构建好牡丹江的水动力模型，然后在其基础上构建牡丹江二维水质模型，以及对这两个模型的适用性验证。当牡丹江水质模型和分布式系统计算构建完成后，就可以整合相关核心技术的实现方法，实现牡丹江水质预警系统的构建。

第 2 章分析了水质模型的发展阶段、定义和分类，讨论了水质模拟预警常用技术方法，包括非机理模型和机理模型。其中非机理模型主要包括数理统计法、灰色系统理论法、神经网络法等，而机理模型主要包括 WASP、EFDC、MIKE、DELFT3D、QUAL2k 等模型工具。本章对比了基于机理模型的水质预警方法和基于非机理模型的水质预警方法各自的优缺点，结合系统研发的具体要求，考虑到 EFDC 的如下一些优点：1）EFDC 软件是美国 EPA 首推的地表水模拟软件，在水动力模拟和水质模拟方面的准确性在学术界有一定的认可度；2）文件的输入方式比较容易理解，更好地耦合水动力、水质和泥沙模块，不同模型接口间的转换过程可以省略；3）EFDC 本身是完

全开源的软件，有利于模型与 GIS 功能组件的集成和二次开发，有利于分布式模拟预警系统的开发。因此，本研究选择了 EFDC 作为研发牡丹江水质模拟预警系统的模型构建工具。本章在最后介绍了该模型构建工具中的部分基本概念、控制方程和数值求解方法。

牡丹江水质模拟模型构建与参数率定研究是预警系统研究的基础，需要结合研究区域的具体情况。比如气象情况、水文情况、河道地形高程和污染本底值等。因此，第 3 章分析了牡丹江研究区概况和水质现状，对牡丹江干流 2000—2014 年各类水质进行统计，在此基础上对各类水质所占比例的年际变化趋势进行分析。除 2008 年以Ⅲ类水为主外，牡丹江流域其他年份的水质类别以Ⅳ类水为主。2009—2013 年，Ⅲ类水比例有所提高，占到了三分之一以上，水环境质量有所提高。2013—2014 年，流域水环境质量又有所下降，特别是 2013 年，Ⅳ类水质所占比例达 85.71%，为 15 年来的最高值；2014 年Ⅴ类水所占比例也达到了 16.67%，同时出现了劣Ⅴ类水质。根据河道地形资料及水质监测资料构建了二维水动力水质模型，模型根据江底高程由西阁断面的 245.7m 降至柴河大桥断面的 215.1m。根据牡丹江干流城市段实际地形，采用正交曲线网格对水体进行剖分，将水体划分为 4207 个单元格，其网格矩阵为 864 行×5 列，单元格尺度介于 24×43.6m—176.9×241.3m 之间。模型流量边界条件包括西阁断面上游来水流量、柴河大桥断面向下游泄水量、海浪河汇入干流流量以及沿江排污口污水排放量。由于该河段内仅在海浪断面处设有水文监测站—牡丹江水文二站，而西阁断面和柴河大桥断面均为水质监测站，西阁断面和海浪河的来水流量根据牡丹江水文二站的流量实测资料进行推算。按照西阁断面上游控制流域面积和海浪河子流域控制面积所占比例，将牡丹江水文二站实测流量资料分配到西阁断面和海浪河入口处。其中，西阁断面流量为牡丹江水文二站流量的 2/3，海浪河入牡丹江流量为牡丹江水文二站流量的 1/3。对于柴河大桥边界条件，为防止模型计算溢出而造成运行终止，该断面采用开边界条件，即水位条件。其水位数据通过牡丹江水文二站实测水位减去水文站所在断面河底高程与柴河大桥断面河底高程之差计算得出。干流排污口排污流量为实测流量数据。浓度边界条件包括西阁断面上游来水污染物浓度、海浪河来水污染物浓度以及沿江各排污口污染物浓度。模型验证断面包括温春大桥、海浪、江滨大桥以及柴河大桥 4 个水质监测断面。通常情况下，各水质监测断面于每年 1、2、5、6、7、8、9、10 月份的月初进行监测。其中，1、2 月份代表冰封期水质，其余月

份代表非冰封期水质（其中 7、8、9 三个月是丰水期，5、6、10 三个月是平水期）。模型构建范围内包括 11 个主要排污口，另外两个排污口不在模型构建范围内。根据实测资料，这 11 个排污口无连续流量监测数据，仅有全年排污总量数据，因此，排污口的流量采用恒定值。对于污染物浓度，每个排污口每季度监测一次，因此浓度边界条件采用实测值。这 11 个排污口每年共排入牡丹江干流约 6000 万立方米污水，部分排污口污染物浓度较高，已经超过了国家规定的污水排放标准，对河流形成了污染。

之后，第 3 章对冰期和非冰期条件下模型的水平扩散系数、垂向紊动粘滞系数和垂向紊动扩散系数、COD 和 NH_3N 综合衰减速率、河床糙率等进行了率定和验证。参数的确定采用经验值法和试算法相结合，即首先采用已有研究成果率定好的经验参数，然后对经验参数进行调整，使模型模拟结果不断接近实测值，最后确定满足要求的模型参数。为保证模型计算收敛，计算步长为常数 6s。经测试计算，非冰封期内 COD 与 NH_3N 的衰减速率分别为 0.03/day 和 0.05/day；冰封期内，COD 与 NH_3N 的衰减速率分别为 0.01/day 和 0.02/day。将不同时段内模拟结果与实测值进行比较并对其误差进行统计分析。牡丹江干流 COD 实测值与模拟值结果对比，其中 COD 模拟值与实测值的变化趋势能够较好的吻合。整体来看，四个验证断面中，柴河大桥断面的平均相对误差最大，为 18.43%；温春大桥断面平均相对误差最小，为 5.86%。从不同模拟时期来看，四个验证断面非冰期的模拟效果均比冰期的好。其中，温春大桥和江滨大桥冰期与非冰期的模拟误差基本接近，而海浪和柴河大桥冰期与非冰期的模拟误差较大，特别是柴河大桥冰期的平均相对误差明显比非冰期的高出许多。造成冰期模拟效果差于非冰期模拟效果的主要原因是冰期内的实测值较少。NH_3N 模拟值也可以反映实际的变化情况，四个验证断面中，柴河大桥断面的平均相对误差最大，为 39.58%；温春大桥断面平均相对误差最小，为 14.88%。从不同模拟时期来看，与 COD 模拟效果不同，四个验证断面非冰期的模拟效果各有差异。其中，温春大桥和江滨大桥冰期模拟效果优于非冰期的模拟效果，而海浪和柴河大桥非冰期模拟效果比冰期的好。总体而言，该模型用于牡丹江干流的 NH_3N 模拟也是可行的，但模拟精度没有 COD 的模拟效果好。造成这一现象的主要原因是 COD 的污染源主要来至工业排污，其污染物浓度和排污量容易监测和控制，模型浓度边界条件较好；而 NH_3N 的污染源主要来至流域内的面源污染和生活污水，因缺少面源污染的监测资料，导致 NH_3N 浓度边界条件准确度下降，进而影响到模型模拟精度。

第 3 章最后分析了模型的研究结果，表明：1）牡丹江水质模型应用于牡丹江干

流城市段 COD 和 NH_3N 水质指标模拟，其模拟精度较高。模拟结果能够客观的反应 COD 和 NH_3N 在该江段中冰期和非冰期时段内的输移过程，将该模型应用于牡丹江干流的水质预测与预报是可行的；2）冰封期内 COD 和 NH_3N 衰减速率要低于非冰封期内的衰减速率。影响衰减速率大小的主要因素为气温、上游来水量以及冰层覆盖。因此，要模拟牡丹江河道内污染物的迁移变化过程，需分冰期和非冰期分别进行模拟；3）糙率是影响河流水动力过程的一个重要参数。非冰封期内，河道的糙率仅考虑河床的糙率，而当河流处于冰封状态时，河道的糙率除河床糙率外还要考虑冰盖阻力的影响，要分别率定冰封期和非冰封期的糙率；4）与中国其他河流的衰减速率相比，牡丹江的 COD 和 NH_3N 衰减速率处于较低水平，这可能与牡丹江地处北方寒区，年平均气温较低有关。

经验证后的牡丹江水质模型可以和 GIS 耦合集成，构建水质模拟预警系统。接下来的章节对水质模型与 GIS 的耦合以及水质模型分布式计算方法进行了研究。论文的第 4 章详细地分析了牡丹江水质模型和 GIS 在预警系统中的耦合集成方法，根据水质机理模型的计算特征，把一种紧密耦合集成的方法应用到本研究中。由于水质模型的密集性计算特征，考虑到水质时空模拟和水污染应急的实际效率需求，设计了一种分布式的模型计算方法，将模型的计算分布到弹性节点服务器集群中，明确了这种水质模型动态弹性计算的分布式架构。在这种架构下设计了分布式水质预警系统的服务器节点间的通信结构，服务器节点和客户端间通信机制，以及总服务器获取弹性节点状态的方法。另外，考虑到方案文件在服务器弹性节点间以及和客户端间的传输问题，结合现有的分布式文件系统，设计了水质模拟方案的二进制存储管理方法，设计了模型模拟方案分布式共享方式以及在这种方式下的方案文件存储压缩方法，文件共享与模型模拟结果存储方法。这种方式可以显著减少服务器节点间的方案数据传输量，从而在同等的网络条件下可以提高模型分布式计算效率。分布式服务器弹性节点和分布式文件共享是实现牡丹江水质模型分布式计算的重要基础。

在阐述完分布式模型编程后，第 5 章讨论了分布式任务计算的任务调度负载均衡，分析了弹性节点的主要评价指标，包括节点可用存储空间、CPU 占用、网络传输负载、文件存储量和可用内存等。讨论了数据存储的负载均衡计算和弹性节点计算的动态负载均衡调度，对水质模拟方案下载的 Map/Reduce 负载均衡调度也进行了阐述。然后分析了水质模拟的分布式计算，包括为解决单个节点的多模型方案并行计算问题和

Map/Reduce 负载均衡调度策略。弹性节点的水质模拟计算线程池的作用是用来提高单服务器节点的多方案模拟能力。Map/Reduce 负载均衡调度的 Map 操作为弹性节点执行模拟的过程，Reduce 操作为客户端下载模拟结果并展示的过程。为确保最终选择的弹性节点可靠，总服务器在接收到客户端水质模拟计算请求时，依据总服务器收到的信息选择其中部分优先服务器，客户端获取最新的节点正在执行的任务数信息。为提高任务执行的可靠性，在选择最佳节点时，采用循环检查选择最佳节点的方法，处理任务执行节点失败信息。

第 6 章对水质模拟预警系统的逻辑结构和功能框架进行设计，把水质预警系统从逻辑上分为物理层、数据层、应用层和表示层，分别对这些层进行了阐述。在此基础上讨论与预警系统相关的几个关键技术的实现方法，包括水质预警系统数据库、水质评价、水质趋势分析、超标预警预报、常规水质和应急水污染预测预警等，其中常规水质和应急水污染预测预报是系统的核心功能，构建在前述章节构建的水动力水质模型基础之上。作者在该章阐述了水质预警系统的原型，描述了弹性节点管理工具和总服务器管理工具。通过系统新建方案工具，创建情景模拟方案，当系统开始运行方案时，系统后台做的工作依次为：1）以每个执行节点的当前执行模型模拟方案数量为判断依据，找出执行数量最少的最佳节点；2）通过方案 ID 查询出本地方案文件；3）读取出本地方案的数据并传递到最佳执行节点；4）该执行节点接收方案后，存储到相应的目录，并以方案 ID 命名文件进行唯一标识；5）该执行节点将从模板中复制一份牡丹江水质模型文件，并以方案 ID 命名文件夹进行唯一标识；6）该执行节点读出方案数据，并将其替换到复制后的牡丹江水质模型文件中；7）该执行节点尝试在多线程中运行该替换后的牡丹江水质模型；8）若运行成功，则将该方案所在的路径登记到分布式文件系统；9）若在上述第 3）至 7）步中，返回了错误信息，系统将排除该节点，并重新回到第 1）步执行，直到方案运行成功；10）水质预测模型在执行节点运行完成后，该执行节点把模型运行完成的标志文件登记到分布式文件系统；11）下载模拟结果，将通过分布式文件系统查询出该方案的模拟数据存在于何弹性执行节点的何路径下，然后进行相应的下载；12）模拟结果文件下载后，在 GIS 客户端进行读取，并且进行时空动态展示。整个流程实现的示例代码请参看附件一。

第 6 章最后利用部署在阿里云弹性节点集群中的服务器进行性能测试，为了简化测试的复杂度，所有弹性计算节点服务器都如 6.4.1 节中描述的最基本配置一模一样。

当采用单节点方式进行预测模拟时，每个方案独占服务器，总耗时是这些方案的耗时总和，即 4.48 小时。但是采用分布式模型计算方式，总耗时取决于耗时最长的情景模拟方案，因此其总耗时为 1.1579 小时，仅仅约为单节点模式的四分之一，也即分布式情景模拟分析的效率得到了显著的提高。采用分布式计算方式，大大提高了情景模拟的计算效率。同时采用这种方式当情景模拟的方案规划发生变化时，可以按需增加或者减少节点服务器的数量，在模拟预测效率提高的情形下节省了服务器硬件设备的使用成本。如果模拟预测河道范围更大，时段更长，模型网格数更多，分布式模型计算的优势会更加明显。

第 7 章是本论文的最后一章，对全文的研究内容进行总结，分析论文的创新点并且对未来研究的方向进行展望。研究成果归纳如下：1）牡丹江二维水质模拟预警模型：该模型考虑了冰封期和非冰封期水污染在水体中传播的差异性，通过污染指标 COD 和 NH_3N 的模拟预测来验证不同水期水质模型构建所需的对流扩散系数、河床糙率及综合衰减速率，因此可以利用牡丹江有限的断面监测数据，开展溶解氧、总磷、总氮等类似污染因子的水质时空模拟预警研究。2）一种分布式的水质模型计算方法：提出并且构建了水质模拟预警模型的分布式计算架构，实现了水质模型模拟方案的分布式存储、共享和下载，实现了水质模型的动态弹性计算，提高了水质预警系统的分布式动态计算能力和应急处置反应效率。3）分布式水质模拟预警原型系统：该系统可以对污染物在水体中的对流扩散趋势和变化状态进行定量化描述和可视化分析，结合水质评价、水质趋势分析、常规和应急水质模拟预警等功能，实现了水质时空动态分析和预测预警，提升了流域水环境质量精细化监控管理能力。

关键词：预警；水质模拟；水动力模型；二维模拟；分布式计算；情景分析；地理信息系统

② Abstract

The consecutively happening water pollution events in China gave us the alarming that the water pollutions are disastrous, influence of the events is far-reaching with a heavy sacrifice in addition to the enormous economic loss. In order to strengthen capability of water pollution forecasting and early warning, relieve the harm to water ecology, animals and human, reduce economic loss and improve human settlement environmental quality, it is extremely necessary to carry out studies on Simulation and Early Warning System (SEWS) of water quality so that the spatio-temporal variation and its change trend can be accurately simulated, forecasted and evaluated for reasonable actions and policies to administrate the water environment in the country.

Mudan River, an important river in northeastern cold regions of China, flows into Songhua River, which, after merging with Heilong River, continues its way running out of China into Russia. Thus, the quality of water in Mudan River is highly concerned not only locally and regionally but also internationally. Due to long winter season, SEWS of water quality in Mudan River would face many difficulties as a result of ice-covered surface for approximately 5 months in a year. As to the Mudan River, SEWS of water quality is a powerful tool for the river water quality control to guarantee safety of water environment. it is urgently to integrate the GIS techniques with the water quality simulation models to monitor sewage outlets discharge state of rivers especially those outlets with over-loaded drainage and un-reported discharge, to support day-to-day administration to the environment and water quality of the basin, to reinforce water environment monitoring and early warning, to reduce the occurrence of water pollution events, to strengthen capability of water quality forecasting and early warning when pollution occurs, to provide decision information support to administration of water pollution emergency, and finally to reduce the possible economic loss as a result of pollution event when happening. Thus, the great significance can be summarized as the following points to the study of SEWS in Mudan River basin.1) Strengthen the study on river water protection technology

in northern cold region of China; 2) Avoid international discord and achieve goal of national strategic plan; 3) Promote routine monitoring and emergency reaction efficiency of river water quality.

Therefore, objective of the study is to establish a practical SEWS for forecasting and early warning of water quality dynamics in the river so that the spatio-temporal distribution of water quality in both open-water and ice-covered periods can be accurately simulated using the available data on water quality from very limited observation sites to understand the spatial variation of pollutants along the river course and their consequence. Based on the simulation results, not only water quality in the river can be evaluated through table outputs, graphs and visualized maps but also early warning can be launched to the pollution indexes when exceeding the pre-defined criteria and to the potential pollution events when suddenly occurring.

Establishment of the water quality SEWS in Mudan River may encounter several challenges. The first is due to the fact that available observation data on water quality of the river are very limited. It is very difficult to calibrate the required parameters for the modelling system on the basis of the limited observed data. Moreover the available data on water quality from the limited observation sites may not be sufficient to conduct a scientific sound simulation so that the spatio-temporal distribution of the water quality in the entire watercourse can be evaluated and the influence of polluted water from the sewage outlets on the river water quality can be monitored. In order to avoid this disadvantage in our case of Mudan River, we built the SEWS on the basis of the mechanism model EFDC (The Environmental Fluid Dynamics Code) recommended by US EPA (Environmental Protection Agency) to simulate the physical and chemical processes of pollutant diffusion and transport over river course. The second challenge comes from the computation of the modelling system. As a distributed modelling system with thousands of grids to be coincidently calculated for numerical scenario solutions, the computation work of the system would be a burden task that consumes long time and high computational resource of servers. Thus it is very important to find solution to improve the scenario analysis efficiency and visualization of the system for spatio-temporal propagation of water pollution, so that the response time of the simulation with the system can be greatly reduced to the emergency treatment when an accident occurs. We solved this problem through integration with GIS (Geographic Informational System) to visualize the water quality and adoption of distributed models computation, leading to a very high efficiency in different scenario analyses of water quality change in the river. The third challenge is how to accurately forecast the spatio-temporal change of water quality in the river so that a better action can be taken as early as possible to deal with potential water pollution event when occurs to minimize the loss and disaster effects of the accident. Several scenarios have been developed in the study to provide a very good estimation for prediction of water quality dynamics in the river.

Thus the dissertation is structured into 7 chapters. Chapter I outlined the background and significance of the study, reviewed the current progress in cold region modelling for water quality, discussed the international research status of SEWS of water quality. For comparative study of hydrodynamic force and water quality process in ice-covered and open-water periods of Mudan River in China, study which establishes 2D water quality model still needs to be intensified when compared with that in European and American countries. What differences exist in hydrodynamic process and pollutant migration and transformation process of watercourse in ice-covered and open-water periods, what are the reasons causing these differences, what are the main influencing factors or indexes, these questions are worth studying. Hence, simulation study of water quality of rivers in northern cold region of China which takes Mudan River as a representative especially water quality in ice-covered period is of great significance in order to better construct river water quality SEWS services in cold region like Mudan River. In this chapter, the methodology and framework of the study were also presented. Review on relevant literature has to be done to understand the current progress of water quality modelling especially the model development and principles of various models for the modelling. The required data especially water quality monitoring data were collected for analysis. Then we analyzed the methods and techniques for water quality early warning and compared the advantages and disadvantages of the methods and the characteristics of the existing water quality models so that a best one can be selected as the model construction tool to establish the modelling system for simulation and early warning of water quality in Mudan River. Current water quality state of Mudan River needed to be analyzed before constructing water quality model, and then hydrodynamic model of Mudan River was constructed and calibration and verification of parameters of hydrodynamic model was performed. After hydrodynamic model was verified, water quality model would be constructed and parameters calibration and verification of water quality model would be performed. Mudan River water quality model which has been verified is very important for distributed SEWS construction. It's necessary for the SEWS to consider integration of GIS and Mudan River water quality model, distributed framework design, distributed file system, distributed computation and equilibrium scheduling method, and then integrated system framework design of water quality early warning and implementation of relevant core technologies. Early warning of routine water quality and sudden water pollution in core technologies will use verified Mudan River water quality model. They are also the key technologies of the system constructed in this study. Finally we conducted scenario simulation of routine water quality and sudden water pollution in order to demonstrate the constructed system and models.

Chapter II analyzed development phase, definition and classification of water quality model, and introduced the common technology method of water quality simulation and early warning, including mechanism model and non-mechanism model. Non-mechanism model mainly adopted mathematical statistics method, grey system theory, neural network method and so on. However, mechanism model mainly contained CE-QUAL-ICM, WASP, EFDC, MIKE, DELFT3D, QUAL2k and so forth. This chapter compared the advantages and disadvantages among water quality simulation methods of mechanism models and non-mechanism models, and considered features of research objective and advantages of EFDC, then selected EFDC as the model construction tool to establish the water quality model in Mudan River, which was integrated into the GIS environment in view of its following advantages: 1) EFDC is the surface water simulation tool that is strongly recommended by EPA of the United States. Its accuracy in hydrodynamic simulation and water quality simulation is accepted in the academic circles to a certain extent. 2) Document input is easy to understand and can make better coupling of hydrodynamic, water quality and sediment modules. Transformation between ports of different models can be neglected. 3) As a completely open source software, EFDC is not only good for integration of the model and GIS functional components as well as further development, but also conducive to developing a distributed SEWS. To simplify the intensive model computation, improve extensibility of model simulation and reduce hardware resource waste, distributed dynamic elastic computation of water quality model was studied when establishing the SEWS. Finally, this chapter introduced the model concept and theoretical foundation involved by the research.

The construction of water quality model of Mudan River is the basis of SEWS, which needs to think about the specific situation of research objective. Chapter III introduced the watershed and analyzed the water quality status of Mudan River, the observed data from the year 2000 to 2014 was used in the analysis. It shows that except for year 2008 in which grade III was dominated, the rest water grade of Mudan River were grade IV. From 2009 to 2013, grade III portion was higher, taking up over 1/3, indicating the water quality was improved. From 2013 to 2014, the water quality again dropped, especially in 2013, water of grade IV took up a percentage of 85.71%, reaching the peak within the fifteen years. Water of grade V in 2014 also took up 16.67% and water of inferior grade V (+V) occurred as well again after year 2010. According to the topography data and water quality monitoring data of the watercourse, this chapter detailed how to make the river grid system and set up the model boundary condition, a new 2D hydrodynamic water quality model has been established to simulate COD and NH_3N concentrations during the ice-covered and open-water periods for the Mudan River in China. Orthogonal curvilinear grid is employed to divide its water body into 4,207 cells with a scale ranging between 24×43.6m and 176.9×241.3m. In addition, the grid matrix is 864 lines x 5 columns.

Bottom elevation of the simulated river section lowers down from 245.7m on Xige section to 215.1m on Chai River Bridge section. Boundary conditions of the simulated flow include upstream on-coming flow of Xige Section, downstream water discharge from Chai River Bridge Section, flow from Hailang River fed into the trunk stream and quantity of wastewater effluent from sewage discharge outlets along the river. Within the simulated river section, as the only hydrology monitoring station named as S2 Station is established on Hailang Section and water quality monitoring stations are set on Xige and Chai River Bridge sections, the on-coming flows for Xige and Hailang sections can be calculated on the basis of observed flow data in the S2 Station. According to the proportions taken by drainage areas controlled upstream of Xige Section and sub-stream of Hailang River, respectively, observed flow data from the S2 Station are allocated to Xige Section and the mouth of Hailang River. Among them, flow of the former occupies 2/3 of that in S2 Station, while flow of the latter feeding into Mudan River accounts for 1/3 in that of such a station. Concerning boundary conditions of Chai River Bridge, the open boundary condition which is also known as the water level condition is adopted. Corresponding water level data can be acquired by using the observed water level in S2 Station minus the difference between the bottom elevation of the section and such an elevation of the Chai River Bridge section. It should be noted that pollution discharge at the sewage outlets of the trunk stream is observed flow data. Boundary conditions for concentrations incorporate pollutant concentrations in upland water from upstream of Xige Section, in inflows from Hailang River and in all sewage outlets along the river. Four water quality monitoring sections such as Wenchun Bridge, Hailang, Jiangbin Bridge and Chai River Bridge are model verification sections. Monitoring water quality on section are carried out in January, February, May, June, July, August, September and October of each year; among them, January and February represent water qualities in ice-covered period, while the remaining months stand for those in open-water period. Within the trunk stream simulation section, 11 main sewage outlets are covered. According to the measured data, there is no continuous flow monitoring data from these 11 sewage discharge outlets, and only the total pollutant discharge of the whole year is available. Consequently, a constant value is adopted for the flow of those outlets. As the concentration of pollutants is monitored quarterly at each outlet, observed values are employed for boundary conditions of concentrations. Every year, sewage of 60 million m^3 approximately is discharged into the Mudan River trunk stream from those outlets; moreover, pollutants from some outlets are highly concentrated, even above the national sewage discharge standard and contaminated the river.

Moreover, we calibrated and verified the involved parameters, including the dispersion coefficient, the roughness and the comprehensive decay rate. On the basis of our research and analysis, some conclusions can be drawn as follows: 1) Research findings show that the concentration simulation errors of COD in the four sections for model verification range from 5.86% to 18.43%; while for those of NH_3N, are between 14.88% and 39.58%. 2) The decay rate of COD and NH_3N during the ice-covered period is lower than that in the open-water period. According to the research results, in the trunk stream of the Mudan River, the decay rates of COD and NH_3N during the open-water period are 0.03/day and 0.05/day while they are 0.01/day and 0.02/day during the ice-covered period. 3) In this research, the roughness adopted for the ice-covered period was 0.043 and 0.035 for the open-water period. Obtaining favorable simulation effects indicates that these two parameters were selected appropriately. 4) Comparing with the decay rates of other rivers in China, those of COD and NH_3N in the Mudan River are relatively lower. This may be due to the low annual average temperature of this river which is located in the cold north region. Although simulated values of COD and NH_3N display excellent consistency with their observed values, certain errors still exist. It has been found that the main reasons leading to pollutant concentration calculation errors included the following aspects: 1) Error of bottom elevation generalization; 2) Variability of pollution sources discharged into the watercourse from urban storm drainage pipe networks and sewage pipes; 3) Dispersed non-point agricultural source pollution on each side of the trunk stream and relevant measurement difficulties. As a whole, after the analysis and verification of the model, it's satisfied and can be coupled to

GIS to construct SEWS.

Chapter IV analyzed the integration method of water quality model of Mudan River and GIS in early warning system as well as the distributed architecture of dynamic elastic computing of water quality model. According to the calculation characteristics of water quality mechanism model, tightly coupled integration method was applied to this study. Considering the water quality spatio-temporal simulation and actual efficiency requirements of the water pollution emergency, we designed a distributed model computing method, distributed of the model calculation into the elastic node server in the cluster, made clear the dynamic elastic computing distributed architecture of water quality model. The communication method among server nodes of distributed water quality SEWS as well as the communication mechanism between servers and clients are designed under such architecture. In addition, considering the transmission problem of water quality model scheme files and combination of existing distributed file system, we designed the binary storage management methods of water quality simulation scheme file, design the distributed sharing way of model simulation program, file sharing and storage method of the model simulation results. This approach can significantly simplify data transmission among server nodes so as to improve the efficiency of distributed model computation under the same network condition. The distributed server's elastic node and distributed file sharing are important basis to achieve the distributed computation of water quality model.

After the introduction to distributed programming model, chapter V discussed the task-scheduled load balancing of distributed computation, analyzed the main evaluation indexes of elastic nodes, including available storage of nodes, CPU occupation, network transmission load, file storage and available memory, etc. Next, this chapter discussed load balancing computation of data storage and dynamic load-balanced scheduling of elastic node computation and clarified Map/Reduce load-balanced scheduling of water quality simulation scheme. Furthermore, this chapter analyzed the distributed computation of water quality simulation and discussed the water quality simulated computation thread pool and Map/Reduce load-balanced scheduling strategy to solve the parallel computing problems of multi-model schemes of single node. The role of elastic node thread pool is used to improve the simulation capability of single server node scheme. Map operations of Map/Reduce load-balanced scheduling were simulation process performed by the elastic node, Reduce operations were the simulation results and presentation process downloaded by the client. To ensure reliability of the elastic nodes finally selected, when the master server receives a water quality model computing request from the client, it selects some servers based on the available information and send them to the client, so that the client could get the latest information about the number of tasks on the nodes. On this basis, the nodes are ranked in ascending order, and the first one is generally the optimum. In order to improve the reliability of the task execution, when choosing the best nodes, the failure information of task-performing node was processed by the method of cycle checking.

Chapter VI designed the logic structure and function framework of water quality warning system, logically divided water quality warning system into physical layer, data layer, application layer and presentation layer, and respectively elaborated them. After that, this chapter further discussed the realization methods of several key technologies relevant to early warning system including the database of the system, water quality evaluation, analysis on water quality tendency, early warning of exceeding standard of water quality, normal water quality simulation as well as sudden pollution forecasting and early warning. Normal water quality simulation and sudden pollution forecasting and early warning were core function of SEWS based on hydrodynamic water quality model. And the prototype of the water quality early warning system, elastic node management tool and master server management tool were also introduced in this chapter. Finally, this chapter elaborated the application scenarios by constructing both the normal and the sudden pollution forecasting and early warning scenario simulation, and revealed advantage of distributed mode efficiency, and implementation method and application efficiency of distributed early warning system of water quality simulation. We constructed scenario simulation scheme by scheme tool newly constructed by the system, when the system starts operating the scheme, backstage operations of the system are successively: 1) Taking quantity of model simulation schemes currently executed by each execution node as the judgment basis, we find the optimal node with least execution number. 2) Inquire local scheme files through scheme ID. 3) Read data of local scheme and transmit to the optimal execution node. 4) After the execution node receives the scheme, it will be stored into corresponding catalogue and conduct unique identification by naming files with scheme ID. 5) This execution node will make one copy of Mudan River water quality model files from template and conduct unique identification by naming file folders with scheme ID. 6) This execution node reads scheme data and substitute it for copied Mudan River water quality model files, 7) This execution node tries to operate this Mudan River water quality model after replacement in multithreading. 8) If the operation is successful, then the route where this scheme is located will be registered into distributed file system. If among previous step 3) to 7), error information is returned, the system will eliminate this node and return to step 1) until the scheme is successfully operated. 10) After execution node operation is finished in water quality simulation model, this execution node will register completed logo files in model operation into distributed file system. 11) Download simulation results, through the distributed file system, we will inquire which route of which elastic execution node the simulation data of the scheme is located, and then we conduct corresponding downloading. 12) After simulation results files are downloaded, we'll conduct reading in GIS client side and conduct spatial-temporal dynamic display. Example codes implemented by the whole process are shown in Appendix I.

The distributed servers used for scenario analysis were deployed in Ali elastic node cluster are used to conduct performance test, in order to simplify complicacy of the test, servers of all elastic compute nodes which are exactly like the most basic configuration described in Section 6.4.1. When single-node mode is used to conduct forecasting simulation, each scheme monopolizes servers, and total time consumption is the sum of time consumptions of these schemes, namely 4.48h. However, when distributed model compute mode is used, total time consumption is decided by the scenario simulation scheme with the longest time consumption, so its total time consumption is 1.1579h. Distributed computation mode greatly improves computing efficiency of scenario simulation. When this mode is adopted and program planning of scenario simulation changes, number of node servers can be increased or reduced as required, which saves usage cost of hardware equipment of servers on the condition that simulation and forecasting efficiency is improved. If simulated and forecasted watercourse scope is broader, time period is longer, and there are more model grids, advantages of distributed model computation will be more obvious.

Chapter VII is the conclusion of the study to summarize the key points and innovation of the study. Scientifically, it can be said that the study has the following three points as innovation: 1) A two-dimensional SEWS on the basis of EFDC was established in the study for forecasting and early warning of water quality dynamics in Mudan River. Different propagations of water pollution in the watercourse of the river for both ice-covered and open-water periods were developed in the model. The convective diffusion coefficients, river bed roughness, and comprehensive decay rate were carefully verified and calibrated for the model using the available data from the limited observation sites. These verification and calibration ensure the model to be more efficient and accurate in simulation of spatio-temporal dynamics of water quality in the river. A complete list of useful functions in terms of practical application have been developed in the system to meet various goals of water quality administration especially early warning to the exceeding standard of COD and NH_3N , it also can be used to other similar pollution indexes like dissolved oxygen, total phosphorus and total nitrogen in the watercourse of Mudan River. 2) An efficient computational architecture was developed and structured in the study for the system to reduce its burden computation tasks of simulating into an acceptable level for forecasting and early warning of water quality in the river. This improvement enables the system to be efficiently applied to the real operation of water quality simulation in the river. Technically, this also provides support for other similar monitoring and early warning systems for the water environment quality. 3) A prototype of SEWS

was developed. The system allows for a quantitative description and visualized analysis of the convective diffusion trend and the variation state of the pollutants in the watercourse. Using the system we are able to analyze the spatio-temporal dynamic of water pollutants in the river, to conduct forecasting and early warning of the water quality in the river. The main functions of water quality assessments, water quality trend analyses, daily forecasting and early warning of water quality were involved into the system, which can greatly improve the capability of monitoring and management of water environment quality in Mudan River.

Keywords: early warning; hydrodynamic model; water quality; two-dimensional simulation; distributed computation; scenario analysis; GIS

Chapter 1

Introduction

1.1 Research background and significance

1.1.1 Research background

Drinking water safety has been widely concerned in China, due to the severe water quality status in the main river basins of the giant country with over 1/4 of the world total population while its water resource is scare and below the average level of the world. This severe water quality situation is clearly stated in the “2014 China Environmental State Bulletin” issued by Ministry of Environmental Protection of China (MEPC, 2015). As indicated in the Bulletin as well as seen in Figure 1-1, data from monitoring campaigns over the country’s seven big basins (Yangtze River, Yellow River, Pearl River, Songhua River, Huai River, Hai River and Liao River) as well as the river basins in the Zhejiang-Fujian region, the northwest region, and the southwest region showed that, in 2014, the nationally-administrated river cross-sections with grade I water quality (the best) only accounted for 2.8% of the total, 1.0 percentage increase over previous year; grade II for 36.9%, 0.8 percentage decrease; grade III for 31.5%, 0.7 percentage decrease; grade IV for 15.0%, 0.5 percentage increase; grade V for 4.8%, and inferior grade V (+V) for 9.0% (Figure 1-1).

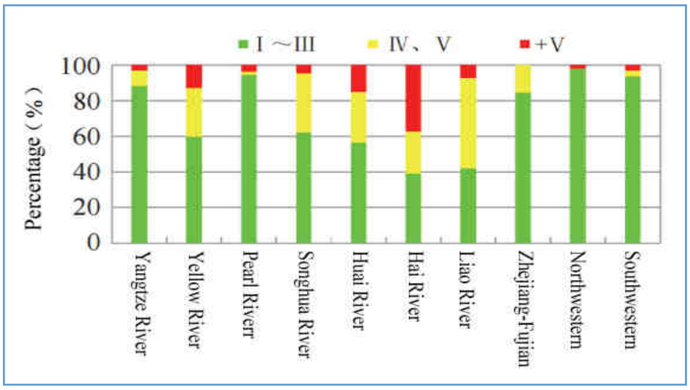


Figure 1-1. Water Quality States of Seven Major Basins and Zhejiang and Fujian Rivers, Northeastern and Southwestern Rivers in 2014.

Therefore, the current overall water quality status is still severe in the rivers of China. This explained why the water pollution events frequently occurred in recent years. There were over 10 water pollution accidents happening in the year of 2014, making the year became the notorious period of water pollution in China. Two serious water pollution accidents consequently occurred in April, 2014 in Lanzhou City, Gansu Province as well as Wuhan City, Hubei Province. Several water pollution accidents happened afterward as a result of heavy-metal contamination, leading to serious injuries and casualties in the country. It was reported that, in December 2014, amoxicillin was detected in the tap water in Nanjing City, Jiangsu Province. Several days afterward, it was uncovered that antibiotics were detected in many big rivers of China. The year 2015 was another year with water pollution events frequently happening in China. Some of the reported environmental pollution events can be listed as follows: Free-captive spiral shells pollution occurred in Jinghang Grand Canal in January. Dead pigs pollution attacked Linwang Stream in Quanzhou, Fujian Province in March. Watercourse pollution was generated by the polluted water discharge of Changyang Mengte Mn-Industry Company in April. Lian River water pollution occurred in Guangdong Province in June. Thousands of square meters of fertile farmland in Chizhou, Anhui Province turned into wasteland. A series of water pollution events like antimony leakage in Gansu Province emerged in November. These frequently-occurred water pollution events have knocked on the alarm to all us that the problem of water safety has been elevated to the level of our survival as a sustainable society and the development of national economy.

The State Council of China (SCC) officially issued “the Action Plan for Water Pollution Prevention and Control” on Aril 2, 2015, which has been nicknamed as “the Ten-Measure Action Plan (TMAP)” by the public. The TMAP sets the national goals of pollution control into three stages: By 2020, water quality all over China will have been obviously improved, and seriously polluted water bodies will be significantly reduced. By 2030, China will work hard to reach an overall improvement to water quality and to preliminarily recover the essential functions of aquatic systems. By the middle of this century, ecological environment quality in China will be completely improved, and ecosystem will have also been fully recovered. The 12th point in Article IV of TMAP clearly declares to “reinforce water environment monitoring and early warning, and strengthen technological support” (SCC, 2015).

The Study of Water Quality Distributed SEWS in Mudan River was therefore proposed under this background to strengthen the capability of water quality monitoring, forecasting and early warning to the basin of Mudan River.

1.1.2 Research significance

The consecutively happening water pollution events in China gave us the alarming that consequence of water pollution events is disastrous and influence of the events is far-reaching with a heavy sacrifice in addition to the enormous economic loss. In order to strengthen capability of water pollution forecasting and early warning, relieve the harm to water ecology, animals and human, and reduce economic loss and improve human settlement environmental quality, it is extremely necessary to carry out studies on water quality SEWS for the main river basins of China so that the spatio-temporal variation and its change trend can be accurately simulated, forecasted and evaluated for reasonable actions and policies to administrate the water quality in the country.

As to the Mudan River, it is urgently to integrate the GIS techniques with the water quality simulation models to monitor sewage outlets discharge state of rivers especially those outlets with over-loaded drainage and un-reported discharge, to support day-to-day administration to the environment and water quality of the basin, to reinforce water environment monitoring and early warning, to reduce the occurrence of water pollution events, to strengthen capability of water quality forecasting and early warning when pollution occurs, to provide decision information support to administration of water pollution emergency, and finally to reduce the possible economic loss as a result of pollution event when happening. Thus, the great significance can be summarized as the following points to the study of SEWS in Mudan River basin.

(1) Strengthen the study on river water protection technology in northern cold region

Mudan River is of representativeness in rivers of northern cold region. Snow and ice cover adds frictional resistance to upper boundary of river water body, changes hydrodynamic characteristics of running water and changes silt transportation process (Debolskaya, 2010). Temperature variation and so on will cause changes of hydrodynamic model parameters, comprehensive pollution attenuation rate and roughness rate, pollutant propagation in river has obvious difference from that in open-water period, so it's necessary to conduct simulation of water quality variation in ice-covered and open-water periods by water quality mathematical model.

Mudan River, the second largest branch of Songhua River, originates from Mudan Ridge of Changbai Mountain in Jilin Province. As it is located in cold region, Mudan River experiences more than 5 months ice-covered period each year. Multiple residential areas and industrial areas are distributed along the bank of mainstream of Mudan River, a large quantity of industrial wastewater and domestic wastewater are discharged into Mudan River, so its water quality is not desirable. Research results show that water body in urban segment of mainstream of Mudan River presents characteristics of organic pollution, and the main standard-exceeding indexes are chemical oxygen demand (COD) and ammonia nitrogen (NH_3N) (Hao et al., 2013). Figure 1-2 shows the location of Mudan River basin and relevant drainage systems.

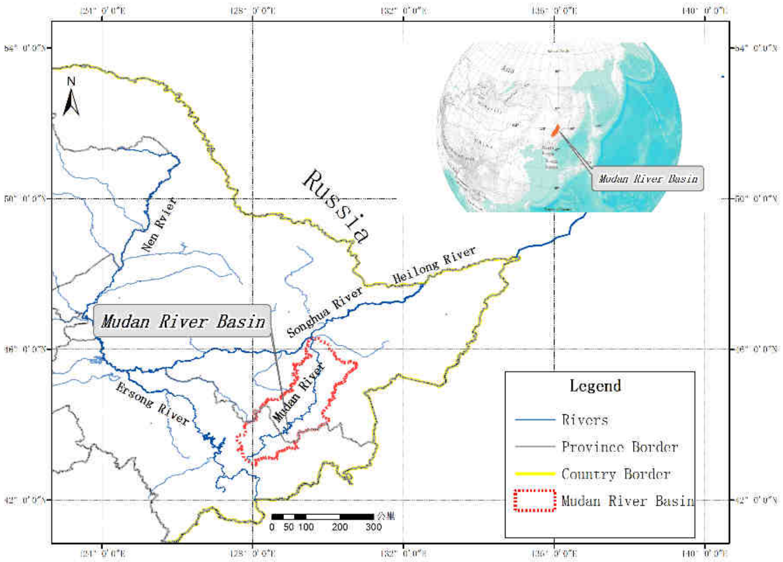


Figure 1-2. Schematic Diagram of Position of Mudan River Basin and River System.

(2) Avoid international discord and achieve goal of national strategic plan

Water pollution in Mudan River may cause international discord. Water pollution in Mudan River basin may be directly fed into mainstream of Songhua River and then further pollute water quality of out-of-border Heilong River (as shown in Figure 1-2). Out-of-border river pollution may trigger environmental diplomacy dispute between China and Russia and cause serious international influence, for example, international influence of the major nitrobenzene pollution event occurring in Songhua River in 2005 was quite intense.

In addition, monitoring management ability in the basin should meet planned requirements in Outline of Advanced Environmental Monitoring and SEWS Construction (2010-2020) issued in December of 2009 and Action Plan of Water Pollution Prevention and Control issued in April of 2015 by Ministry of Environmental Protection—it's necessary to construct advanced environmental SEWS, develop comprehensive analytical tool of monitoring data and early warning information distribution platform, realize visualized expression of forecasting and early warning simulation analysis, improve comprehensive analysis and evaluation level of environmental quality, and promote effective application of monitoring data in programming and planning, policies and regulations, environmental impact assessment, pollution prevention, ecological protection, environmental supervision, etc.”. Therefore, with a view to safeguarding drinking water safety for residents in the whole Mudan River basin centering on 2.8 million population in Mudanjiang City, it’s necessary to pay close attention to perfect and promote water quality monitoring and SEWS study and construction work in Mudan River, however at the moment, there is no similar water quality SEWS in this basin, so construction of water quality SEWS in Mudan River is especially necessary and urgent.

(3) Promote routine monitoring and emergency reaction efficiency of river water quality

Water quality model computation in water quality SEWS is a kind of compute-intensive application, especially under the situation in which simulated river is longer, simulated grids divided are finer, simulated pollution indexes of the model are increased, simulated step length is shortened and simulation interval is lengthened, for model simulation servers of the same configuration, all of these will obviously increase time consumption of water quality model computation, increase time occupation of water quality SEWS, impact standard-exceeding water quality and pollution accident emergency processing efficiencies, hold up formulation of emergency scheme and may cause enormous losses. If traditional high-performance server is purchased, computing power of the server is fixed, while requirements of water quality simulation for hardware are in dynamic change, which will cause huge hardware expenditure and results in unnecessary waste. Advantages of distributed computation: it can conduct dynamic expansion of water quality model computation, realize flexible invocation of distributed multiple simulation schemes, and enhance distributed dynamic flexibility and parallel computation capacity of multiple water quality simulation schemes of SEWS.

1.2 Research progress

1.2.1 Water quality simulation study state in cold regions

Where Mudan River basin is located belongs to cold region in China, and ice-covered river is a common natural phenomenon in cold regions. Ice-covered phenomenon may appear in every winter in regions to the north of northern latitude 30°in China and Tibet Plateau region. Ice -covered probability is greater in higher latitude with longer ice-covered period. In northeastern region of China, ice-covered period of river even reaches more than 5 months. Ice-covered river will generate many adverse effects on production and life of human, such as loss of river shipping functions, ice flood caused by ice jamming or ice dam, frost heaving damage of hydraulic structure, etc. (Manolidis et al., 2014; Fu et al., 2014). In fact, influenced by ice sheet, water quality characteristics, mixed ability, transportation and diffusion properties and so on of ice-covered river dramatically changed.

Some scholars (Chambers et al., 2000; Prowse, 2001; Neto et al., 2007; Martin, 2013) studied DO (Dissolved Oxygen) change rules of ice-covered watercourse. Chambers et al. (2000) adopted 1D steady-state water quality model to analyze the influence of pulp mill and urban pollution discharge in northern Canada river basin in ice-covered period on aquatic organisms, analyzed DO variation under ice of different temperatures in different spatial positions of the watercourse in details, and didn’t use hydrodynamic model during study process to simulate water flow state. Prowse et al. (2001) analyzed and concluded that because of ice cover influence in cold region, variation of topographic form and chemical reaction process changed erosion and sediment process in water, and this change influenced DO variation. Specific influence generated by hydrodynamic force on river water quality was not analyzed in the study which only concluded study results of predecessors, in the meantime, it compared and discussed about the influence of ice-covered river on biological conditions. Neto et al. (2007) enhanced the influence of oxygen content in sewage in river water quality through artificial aeration technique in order to study DO variation in river when pulp mill discharged sewage under ice-covered and open-water conditions. The study used hydrodynamic model and water quality model to analyze river water quality variation in river vertical two dimension (2D) space caused by artificial aeration, and this study had important reference value for analyzing DO variation in water body discharged by sewage treatment plant under ice-covered and open-water conditions. Martin et al. (2013) used CE-QUAL-W2 model to study DO variation in winter of Athabasca River of northern Canada river basin, and it conducted an analysis with a 2D model by combining hydrodynamic and water quality model and calibrated relevant parameters like hydrodynamic force, air temperature, NH_4, nitrate and nitrite, phosphate and floating algae plants, etc. Sugihara et al. (2013) constructed a lake-type water quality model to study eutrophication process in water body in ice-covered period and analyzed variation of nutritive salt and BOD (Biochemical Oxygen Demand) and so on in water with temperature as ice melted, so it could be applied to water quality forecasting and early warning of the similar situation. Similar water quality studies of lake-type ice-covered period had a common characteristic—they didn’t need to take hydrodynamic process in ice-covered period into consideration or use hydrodynamic model to simulate relevant hydrodynamic information during model construction.

There are few achievements of simulation studies in China on water body quality in ice-covered period through hydrodynamic water quality model. Sun et al. (2012) established hydrodynamic water quality coupling 1D model of Songhua River mainstream in open-water and ice-covered periods according to drainage basin characteristics. Firstly, it used actually measured hydrological and water quality data over the years to establish numerical model in open-water period. Major parameters of the model (longitudinal diffusion coefficient, pollutant attenuation coefficient, etc.) were determined with a combined method of field monitoring and model calibration, used monitoring results to analyze applicability of empirical formulas of two longitudinal diffusion coefficients—Fischer and Elder in Songhua River, based on which it modified the model according to water conservancy factors and hydrological monitoring characteristics and then established hydrodynamic water quality model suitable for this region in ice-covered period. In addition, it used this model to reveal water quality characteristics and mixed characteristics in ice-covered period in urban segment in Mudan River. Based on this analysis, it studied the influence of situational measures such as point source pollution reduction, draining exit arrangement, background concentration control, regulation of water-carrying capacity from upstream, etc on water quality on urban segment of Mudan River.

For comparative study of hydrodynamic force and water quality process in ice-covered and open-water periods of the same river in China, study which establishes 2D water quality model still needs to be intensified when compared with that in European and American countries. What differences exist in hydrodynamic process and pollutant migration and transformation process of riverway in ice-covered and open-water periods, what are the reasons causing these differences, what are the main influencing factors or indexes, these questions are worth studying. Hence, simulation study of water quality of rivers in northern cold region of China which takes Mudan River as a representative especially water quality in ice-covered period is of great significance in order to better construct river water quality SEWS services in cold region like Mudan River.

1.2.2 Study of water quality SEWS under GIS support

Water quality forecasting and early warning method constructed in this work was carried out under the support of GIS technology. Since the 60s of the 20th century, with increasing sudden water environmental pollution events, study of water quality early warning methods has drawn extensive attention and brought enormous social and economic benefits. Some rivers which were seriously polluted in the 50s and 60s of the 20th century like Rhine (Diehl et al., 2006) and Danube (UNDP/GEF，2005) in Europe and German Ruhr (Bode et al., 1999), U.S. Mississippi (Bostock et al., 1991) and so on used water quality monitoring and SEWS as well as spatial database technology, mapping technology and spatial visualization technology of GIS to assist in implementing water protection work in basins and obtained significant achievements.

In order to tackle sudden water pollution in Rhine in 1986, International Commission for the Protection of Rhine (ICPR) installed water quality monitoring and SEWS which was beneficial for investigating water pollution events caused by industrial pollution discharge or ship leakage. This system continuously monitored water quality state, so sources of unreported pollution leakage could be deduced and traced through water quality model or other relevant methods (Diehl et al., 2006). This system collaborated with nine countries to jointly protect water safety in Rhine, it also combined data management, spatial query and interactive map technologies, and people from all around the world could acquire data and research report information related to water quality monitoring and pollution events from its website. In December of 2000, EU member countries implemented EU Water Framework Directive (WFD) which aimed at providing a water environmental protection framework for Europe, EU countries carried out extensive researches on water protection technology under this framework in main European rivers like Rhine, Seine and Danube. Under this framework, a specialized GIS working group (GIS-WG) was founded in 2009, aiming at making better usage of GIS to provide services for water environment protection. This working group carried out a series of research work including spatial basic data, water quality monitoring network, data model and management system research and development, etc. (European Commission, 2015).

Grayman et al. (2000) introduced water quality SEWS developed in U.S Ohio River, and this system consisted of three parts: 1) river water pollutant status analysis module; 2) propagation path of pollutants in water body and pollutant concentration distribution analysis module; 3) information transmission mechanism analysis module of pollutant propagating and diffusing in river during pollutant leakage. This system applied WASP model to analysis of water pollution, in the meantime, used ArcView to process data from water quality monitoring station and water pollution and conduct visualized analysis. Since "9.11" event, U.S. has worried that drinking water may become an attack objective of terrorism crime. Since the end of year 2002, U.S. Honolulu Water Supply Association has commissioned Hawaii University to carry out research program of SEWS of drinking water pollution caused by terrorist attack, vicious poisoning of criminals or accidental water pollution. U.S. conducted vulnerability evaluation and formulated countermeasures for water supply systems of over 8,000 water sources all across the nation, enhanced pollution risk management. State governments, environmental protection agencies and maritime sectors and so on along the rivers jointly formulated practical and concrete countermeasures for sudden water pollutions. SEWS in U.S. water source could provide monitoring data feedback, real time data import and emergency processing decisions, and develop a decision-making platform for water supply departments and related workers to process sudden water pollution accidents (Gulliek，2004). GIS functions provided powerful support for data acquisition, data processing, spatial information query and map visualization. Peng et al. (2010) studied integrating method of GIS technology and WASP model, constructed hydrodynamic analysis system which was applied to water quality analysis in Charles River basin of U.S. Massachusetts in order to demonstrate integration process, but operation of this model after integrating with GIS was in loose coupling mode.

As a whole, European and American countries have an early starting in using GIS technology to study water quality SEWS in a relatively advanced way. At present, main rivers in European and American countries which involve sewage outlets and risk pollution source management have been constructed with SEWS and achieved good application by combining data management ability, drawing ability and visualized expression functions of GIS. However, further study is still necessary for constructing advanced SEWS by combining GIS and distributed computation technology.

In the middle of the 90s of the last century, China started researches on environmental SEWS and what was representative was research and application of environmental early warning conducted by Chen Guojie et al. and they proposed the concepts of status early warning and tendency early warning. Water quality SEWS rapidly developed and became an important constituent of modern water quality management and control. As occurrence and development of water pollution accidents was of uncertainty in spatial-temporal aspect and pollution risk sources, and according to timeliness of pollution accident response and processing, and the principle of minimal loss, Ding et al. (2003) adopted the method of flexible arrangement of information on accident spot to analyze basic situation of pollution accident and realize simulation of influence of sudden river water pollutions. They combined GIS and water pollution model technology and developed water pollution accident simulation subsystem suitable for water environment decision-making in Three Gorges which could reflect water pollution state caused by pollution accidents and its spatial-temporal variation process.

After Songhua River serious water pollution accident occurred in 2005, Chinese scholars paid more attention to study of SEWS. The system in Liaohe River basin consisted of water quality information acquisition module, water environment information query module, information transmission and network system module and decision-making supporting module (Li et al., 2007). Chen et al. (2010) adopted component-based GIS technology, based on ArcEngine SDK and combining business requirement of sudden water pollution accident management system for water quality model, discussed about main contents and methods of constructing sudden water pollution emergency response system based on GIS technology under the support of spatial database, thus providing technical support for water resource management and reducing damage caused by water pollution and so on. Li et al. (2008) carried out a study on Qiantang River water quality SEWS, under Visual Studio.Net 2005 environment, they adopted MapX components of Mapinfo and C#.net to conduct secondary development, thus realizing forecasting and sudden pollution accident simulation of pollutant migration and diffusion. Wu et al. (2009) proposed an engineering model suitable for river sudden water pollution emergency treatment, adopted four-point implicit difference scheme to conduct numerical solution of the model and realized real-time visualization of pollution simulation results on ArcGIS platform. Through real cases, they verified effectiveness and reasonability of application of this model in public safety emergency response platform, but water quality model hadn't been integrated with GIS. Hou (2010) established EWS for trans-boundary major pollution accident based on GIS. Overall design of the system was divided into three parts: early warning map composition, map operating control and data management and maintenance. Then he used ArcMap to complete composition of early warning thematic maps. Afterwards, he used customizing functions of ArcIMS network publish components, designed and implemented geospatial database of transboundary major water pollution accidents, and then used ArcSDE as middleware to realize transmission of vector data in Oracle database. Huang Rui et al. (2013), taking Suzi River feeding into a big reservoir in Northeastern China as an example, carried out key technology research for basic environmental information management, pollution source trace, accident early warning simulation, accident emergency treatment, etc., constructed hydrodynamic and water quality mathematic model according to characteristics of Suzi River basin, studied and established pollution source trace technology, adopted Microsoft Visual basic (VB) and MapX to realize spatial-temporal dynamic visualization of water quality variation in the process of water pollution accident, which could provide technical support for emergency treatment of sudden water pollution accidents in Suzi River basin. It can be known from these literatures that models computation in the system is single-server computation mode which can't realize dynamic invocation of distributed server resources. Computation capacity of parallel multi-model simulation scheme was not sufficient, needing further research.

Effect of GIS technology in water quality SEWS is mainly represented in three aspects, firstly, providing data foundation related to water environment analysis: integrating capacity of spatial information and relevant attribute data, data editing and processing ability and database establishing, maintenance capacity which are exclusive to GIS technology can provide data support for monitoring and early warning of water quality; secondly, providing water environment spatial analysis and visualization ability: analysis capacity of GIS based on spatial location is what other systems don't have, and it can realize multi-dimensional visualized expression of relevant data in early, middle and later stages of spatial simulation analysis; spatial analysis and visualization ability of GIS can be integrated with traditional hydrodynamic model and water quality model in order to improve decision-making supporting ability of SEWS; thirdly, providing functional development support for water environment management or smart decision-making support system: through decades of development, GIS technology has had functional development and supporting capacity in mobile client, desktop client , server client and cloud server client, so all kinds of water environment management information systems, decision-making supporting system, information issuing and sharing systems, smart watershed management systems and so on can be integrated and developed on the basis of GIS functional components or software platforms.

1.2.3 Distributed computation of water quality simulation

General definition of distributed computation is a collaborative computing system which conducts communications only through message passing and in which a hardware or software functional module is distributed on network computer. This distributed system covers all software and hardware systems which effectively deploy networked computer. Computer network in real life has been ubiquitous, and campus network, in-vehicle network, mobile telephone network, company network and so on have similar characteristics. Distributed computing system is right constructed on the basis of these networks, and it realizes resource storage computation and sharing through network. In narrow sense, distributed computing system consists of many components which realize specific functions through interaction among distributed networks (Chambers, 1984). Distributed system is a computing system in which multiple interconnected collaborative processing resources participate, it can cooperatively execute one or multiple common tasks under control of the whole system while not depending on specifically centralized procedure, data or hardware. These resources can not only be physical neighboring but also be spatially dispersed in location.

Distributed computation system is constructed on the basis of network. In the last dozen years, computer network has rapidly developed both in hardware and software, especially emergence and rapid development of cloud computation significantly promotes distributed computing technology. At the moment, distributed system has been ubiquitous. Network nodes participating in distributed computation use internal storage and magnetic disk as storage resources and CPU time slice as computing resources to construct an independent node in distributed system. These distributed nodes provide means of shared resources and shared data and they are critically important in this informationalized era. Among many key development orientations of current computer realm, smart computing, private cloud computing or public cloud computing involve core theoretical problems of distributed computation. Involved actual problems include how to conduct resource distribution, how to communicate, how to conduct data storing and data sharing, etc.

Although distributed system is very important, it’s not easy to design effective algorithm on it for water quality model. This model distributed computation is a computation of single node with one to multiple processors or multiple nodes with multiple processors. Considering intensive-type computation characteristics of water quality model, this work uses this technology to realize distributed computation of water quality model, while this model is relatively independent, and other relevant research achievements haven't realized this function.

1.3 Objective and methodology of the study

1.3.1 Objective of the study

Objective of the study is to establish a practical hydrodynamic modelling system for simulation and early warning of water quality dynamics in the river so that the spatio-temporal distribution of water quality in both open-water and ice-covered periods can be accurately simulated with high efficiency using the available data on water quality from very limited observation sites to understand the spatial variation of pollutants along the river course and their consequence. Based on the simulation results, not only water quality in the river can be evaluated through table outputs, graphs and visualized maps but also early warning can be launched to the pollution indexes when exceeding the pre-defined criteria and to the potential pollution accidents when suddenly occurring.

1.3.2 Methodology and framework of the study

Three critical problems need to be solved in establishing the SEWS for water quality monitoring in Mudan River. Firstly, in order to establish the system for water quality 2 dimension (2D) modelling in Mudan River. A good hydrodynamic model tool has to be selected according to the reality of Mudan River and the research requirements. The required key parameters have to be calibrated for the model in both the ice-covered and open-water periods. So we faced the following challenges: what kind of water quality evaluates model should be adopted; how to get the hydrological parameters used by the model; what conditions need to be satisfied to establish the model; how to calibrate and validate the parameters of the model; whether the data in ice-covered and open-water periods meets the requirements of the study.

Secondly, an efficient method has to be developed for the burden tasks of distributed computation involved in implementing water quality simulation by SEWS. This needs to solve how to integrate the model and GIS in distributed architecture, how to design the distributed architecture, how to realize the communication between servers and clients, how to deal with the data transmission, storage and performance optimization issues.

Finally, implementation method of key technologies of water quality SEWS under technical support of GIS, expecting to construct a relatively complete water quality SEWS with dynamic expansions and computation ability, realize simulation and early warning of daily water quality and emergency water pollution of Mudan River in ice-covered and open-water periods, and provide technical support for water environment safety of typical rivers in northern cold region of China. So we should solve how to implement the main modules involved by SEWS.

According to the above philosophies. Figure 1-3 showed the overall framework of this study. Review on relevant literature has to be done to understand the current progress of water quality modelling especially the model development and principles of various models for the modelling. The required data especially water quality monitoring data were collected for analysis. Then we analyzed the methods and techniques for water quality early warning and compared the advantages and disadvantages of the methods and the characteristics of the existing water quality models so that a best one can be selected as the core to establish the modelling system for simulation and early warning of water quality in Mudan River. Current water quality state of Mudan River needed to be analyzed before constructing water quality model, and then hydrodynamic model of Mudan River was constructed and calibration and verification of parameters of hydrodynamic model was performed. After hydrodynamic model was verified, water quality model would be constructed and parameters calibration and verification of water quality model would be performed. Mudan River water quality model which has been verified is very important for distributed SEWS construction.

It's necessary for distributed SEWS to consider integration of GIS and Mudan River water quality model, distributed framework design, distributed file system, distributed computation and equilibrium scheduling method, and then integrated system framework design of water quality early warning and implementation of relevant core technologies. Early warning of routine water quality and sudden water pollution in core technologies will use verified Mudan River water quality model. They are also the key technologies of the system constructed in this study. Finally we conducted scenario simulation of routine water quality and sudden water pollution in order to demonstrate the constructed system and models.

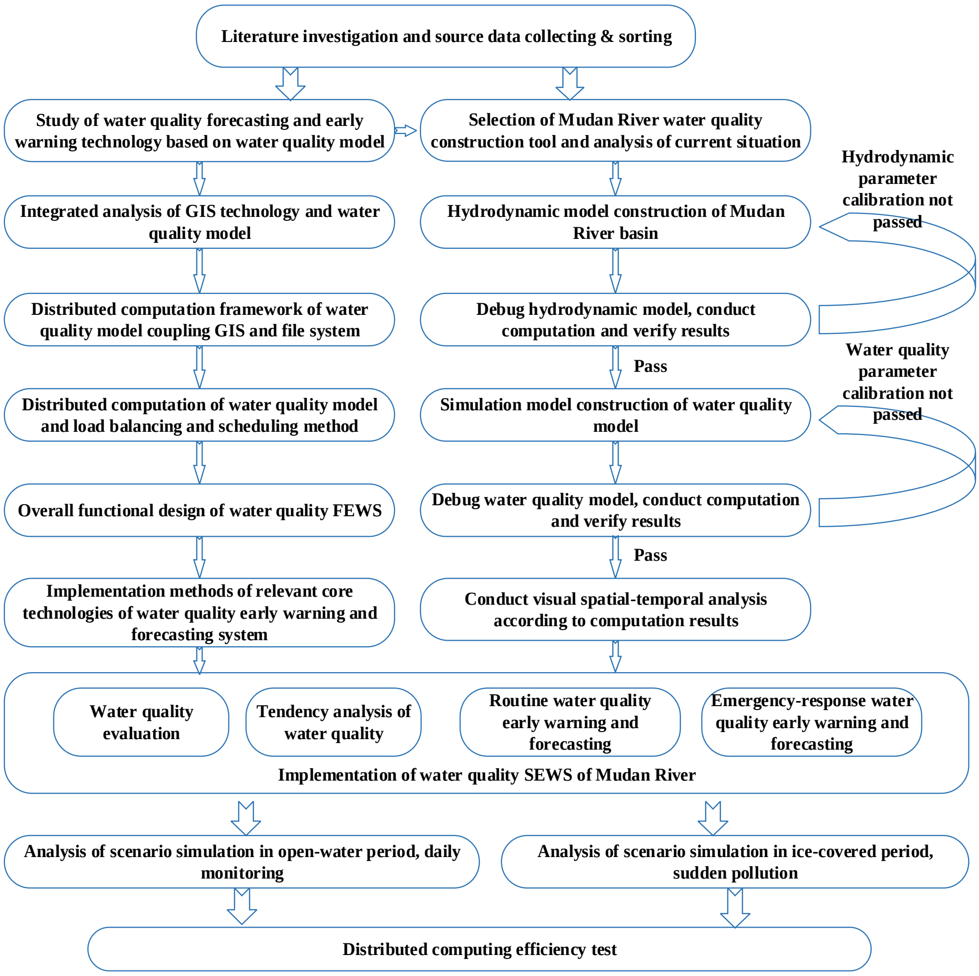


Figure 1-3. Framework of the research.

Chapter 2

Forecast and warning technology based on water quality model

2.1 Introduction to surface water quality model

Water quality model can be effective tools to simulate and predict pollutant transport in water environment, which can contribute to saving the cost of labors and materials for a lot of chemical experiments. Moreover, it is inaccessible for on-site experiments in some cases due to special environmental limitation. Therefore, water quality model become an important tool to analyze water environmental pollution and behaviors of pollutants in water environment (Wang et al., 2013).Surface water quality models have undergone a long period of development since Streeter and Phelps built the first water quality model (S-P model) to control river pollution in Ohio State of the U.S. (Streeter, et al. 1925). Surface water quality models have made a big progress from single factor of water quality to multi-factors of water quality, from steady-state model to dynamic model, from point source model to the coupling model of point and nonpoint sources, and from zero-dimensional mode to one-dimensional, two-dimensional, and three-dimensional models (Wang et al., 2011;Xu et al., 2003). More than 100 surface water quality models have been developed up to now. However, each surface water quality model has its own constraint conditions (Burn et al., 1985). Therefore, water quality models still need to be further studied to overcome the shortcomings of these current models. Generally, the surface water quality models have undergone three important stages since 1925 to now (Wang et al., 2013).

2.1.1 Development of surface water quality model

(1)The primary stage (1925–1965)

Water quality of water body has received much more attention at this stage. The water quality models emphasized the interactions among different components of water quality in river systems which were affected by living and industrial point source pollution (Wang et al., 2004; Rinaldi et al., 1978). Like hydrodynamic transmission, sediment oxygen demand and algal photosynthesis and respiration were considered as external inputs, whereas the nonpoint source pollution was just taken into account as the background load (Mujumdar et al.; 2004; Riffat, 2012). The simple BOD-DO bilinear system model was developed and achieved a success in water quality prediction, and the 1D model was implemented at solve pollution issues in rivers and estuaries (Burn et al., 1985). And then most researchers modified and further developed the Streeter-Phelps model (S-P model). Such as Howland et al. (1949) believed that BOD could be reduced without oxygen consumption due to sediment deposition and flocculation, and the decay rate was proportional to the number of remained BOD; thus, the flocculation coefficient was adopted in the steady-state S-P model to distinguish the two BOD removal methods. O'Connor et al. (1967) divided BOD parameter into carbonized BOD and nitrified BOD and added the effects of dispersion on the basis of the equation. Dobbins and Camp (Dobbins, 1964; Camp, 1963) added two coefficients, including the changing rate of BOD caused by sediment release and surface runoff as well as the changing rate of DO controlled by algal photosynthesis and respiration, to Thomas's equation.

(2) The improving stage (1965–1995)

Water quality models started to pay attention to multiple environmental media and considered distributions and transformations of pollutants in environmental media (e.g. air, water and soil) comprehensively. Multi-media eco-environment integrated model was exploited. Water quality model developed from one-dimensional steady-state ones into multi-dimensional dynamical ones in which hydrodynamic models and sediment models were introduced. Water quality models in this development phase conformed better to practical situations.

From 1965 to 1970, water quality models were classified as six linear systems and made a rapid progress based on further studies on multi-dimensional coefficient estimation of BOD-DO models. The one-dimensional model was updated to a two-dimensional one which was applied to water quality simulation of lakes and gulfs (Welander, 1968; Gough, 1969). Nonlinear system models were developed during the period from 1970 to 1975 (Yih et al., 1975). These models included the N and P cycling system, phytoplankton and zooplankton system and focused on the relationships between biological growing rate and nutrients, sunlight and temperature, and phytoplankton and the growing rate of zooplankton (Gough, 1969; Yih et al., 1975; Riffat et al., 2012;). The finite difference method and finite element method were applied to these water quality models due to the previous nonlinear relationships and they were simulated using one- or two-dimensional models. After 1975, the number of state variables in the models increased extremely, and the three-dimensional models were developed at this stage, and the hydrodynamics and the influences of sediments were introduced to water quality models (Wolanski et al., 1992; Zheleznyak et al., 1992). Meanwhile, water quality models were combined with watershed models to consider nonpoint source pollution input as a variable (Timand et al., 1994; Hunsaker et al., 1995). The effects of sediments were dealt with inner interaction processes of the models (Timand et al., 1994); so the sediment fluxes could change accordingly under different input conditions. Therefore, the water quality management policies were significantly improved due to more constraint conditions and nonpoint source pollution simulation of watershed range. The representative model tools including QUAL (Grenney et al., 1978; Brown et al., 1987), MIKE11 (Danish Hydraulics Institute, 1993), and WASP (Ambrose et al., 1988; Ambrose et al., 1993) were developed and used at this stage. At the same time, the one-dimensional OTIS model developed by USGS was also applied to water quality simulation (Bencala et al., 1983; Valett, et al., 1996).

(3) The deepening stage (after 1995)

Water quality models began to cooperate with nonpoint source models and nonpoint pollution sources could access to initial input of the model. Meanwhile, various new technologies and methods, such as fuzzy mathematics, stochastic mathematics, expert system and artificial neural network, were applied to water quality model study. GIS began to make organic coupling with water quality models, forming a relatively complete river water quality management system.

Nonpoint source pollution has been reduced due to strong control in developed countries. However, the dry and wet atmospheric deposition such as organic compounds, heavy metals, and nitrogen compounds showed increasing effects on water quality of rivers (Golomb et al., 1997; Poor et al., 2001; Morselli et al., 2003). Although nutrients and toxic chemical materials depositing to water surface have been included in model framework, these materials not only deposited directly on water surface but also they can be deposited on the land surface of a watershed and sequentially transferred to water body (Esterby,1996; Mason,1997), which has been an important pollutant source. From the viewpoint of management demands, an air pollution model has to be developed to adopt this process in the model, indicating that the static or kinetic atmospheric deposition should be related to a given watershed (Bai et al., 2011).Therefore, at this stage, some air pollution models were integrated to water quality models to evaluate directly the contribution of atmospheric pollutant deposition (Esterby, 1996).With the exception of the typical model tools such as QUAL2K (Fang et al., 2008), WASP6 (Artioli et al., 2005), QUASAR (Whitehead et al., 1997; Sincock et al., 2002), SWAT (Grizzetti et al., 2003), and MIKE21 (Danish Hydraulic Institute, 1996) and MIKE31(Danish Hydraulic Institute,1996), other water quality models have also been developed to simulate complicated water environmental conditions. For example, Whitehead et al. (1998) developed a semi-distributed integrated nitrogen model INCA based on the effects of atmospheric and soil N inputs, land uses, and hydrology. More recently, Fan et al. (Fan et al., 2009) integrated QUAL2K water quality model and HEC-RAS model to simulate the impact of tidal effects on water quality simulation. For the integration of point and nonpoint sources, the Environmental Protection Agency of U.S. (USEPA) developed a multi-purpose environmental analysis system BASINS, which makes it possible to assess quickly large amounts of point and nonpoint source. Meanwhile, USEPA also listed EFDC model as a tool for water quality management. Among the above-mentioned surface water quality models, these model tools including Streeter-Phelps, QUASAR, QUAL, WASP, CE-QUAL-W2, BASINS, MIKE, and EFDC were widely applied worldwide (Fan et al., 2009; Morley, 2007; Hamrick, 1992). Recently, Kannel et al. (2011) concluded that these public domain models (e.g., QUAL2EU, WASP7, and QUASAR) are the most suitable for simulating DO along rivers. Generally, most developed countries (especially the U.S. or European countries) have developed better and advanced surface water quality models (Danish Hydraulics Institute, 1993; Danish Hydraulic Institute, 1996; USEPA, 1999). Some surface water quality models have also been established in some universities or institutes of China over the past years, but these models were still not widely utilized like MIKE, EFDC, and WASP (Zhang et al., 2009; Wang et al., 2009; Wang et al., 2013).

Functions of water quality models had been enhanced. They paid more attention to usability and simulation accuracy and were further integrated with third systems, including GIS, remote sensing (RS) and decision support system (DSS), to construct a more robust comprehensive river decision management system. Owing to the development of IT, diverse application systems with different software frameworks began to be developed.

With economic, social and scientific developments, people propose higher requirements on water environmental quality and are facing with increasingly complicated water environmental issues. It is necessary to combine mathematical approaches with water environmental chemistry and hydraulics to establish a hydrodynamic and water quality mathematical model for quantitative study on variations of the water environment, chemical and ecological elements, so as to serve eco-environment protection better. Due to decentralized research works and single applications in early period, China hasn't developed any mature water quality model software. Designers and engineers prefer to purchase commercial software from European and American countries, which leads to great waste of funds and hinders development of China's mathematical model of the water environment.

On the whole, European and American countries studied water quality model earlier than China. They have established numerous water quality models, developed corresponding software and applied them widely to water environment planning and management. China's river water quality model needs to be further studied (Song, 2013).

2.1.2 Definitions and classification of river water quality models

SEWS of river water quality based on water quality models has been developed for over 80 years. Different scholars or institutions gave different definitions to water quality model, because they focused on different fields or applications. For instance, USEPA emphasized practical application of water quality models, while Xie and Luo et al. focused on the mathematical relationship when pollutants migrate and transform in the water environment (Song, 2013). Song et al. (2008) defined water quality model as a mathematical equation that is used to describe temporal-spatial migration and transformation law of pollutants in water environments as well as relations among different influencing factors. They viewed water quality model as an important tool for water pollution management and planning.

General procedures of water quality model construction and water quality simulation and early warning include: collecting relevant data of hydrology and water quality of monitoring and meteorological observation to solve water environment problems; on the basis of physical, chemical, biological changes of simulating prediction pollutants and water quality, constructing mathematical equations (model structure) of simulating prediction pollution factors-related factors relation, Solution of the mathematical equations relying on some calibrated parameters (e.g., degradation coefficient, riverbed roughness and so on), namely calibration of model parameters; verifying the ability of the model simulating and predicting water quality state using observed data unused by calibration of model parameters, i.e., verification of model. If calibration and verification achieved requirements, it'll be applied to predicting and analyzing water quality.

Early warning models of river water quality can generally be divided into two types, namely, mechanism model and non-mechanism model (Wu, 2005). Mechanism model establishes a mathematical model mainly according to migration and diffusion law of pollutants in water. It gives quantitative descriptions of convective diffusion trends and changing states of pollutants in water. Mechanism water quality models are generally inferred from equation of motion, equation of continuity and equation of energy according to conservations of energy and mass. At present, common river water quality simulation models include CE-QUAL-ICM (Cole et al., 1995), WASP (Ambrose et al., 1993), EFDC, MIKE (Danish Hydraulic Institute, 2002; 2004; 2005), Delft3d (Delft3D-FLOW, 2006) and QUAL2k (Chapra et al., 2012). Non-mechanism model mainly establishes a mathematical model to predict variation trends of water pollutants by mathematical statistical analysis, grey system theory and other data mining methods. Non-mechanism water quality models mainly focus on a certain water quality system and are characteristic of simple prediction. Existing non-mechanism water quality models include mathematical statistic method, grey system theory, neural network, etc. (Song, 2013).

2.1.3 Comparison among non-mechanism models for water quality

Comparison of major non-mechanism models for water quality are listed in Table 2-1.

Table 2-1. Comparison of major non-mechanism models.

Simulation methods	Theoretical basis	Data demand	Advantages	disadvantages
mathematical statistics method	Theory of probability and mathematical statistics	Rich water quality monitoring data	Only require rich water quality observed data	Complicated statistics under multiple elements
Grey system theory	Grey system theory	Low data demand	Unlimited by observed data , simple	Limited simulation accuracy
Neural network method	Neural network	rich train samples	High-dimensional features high adaptation	Poor effect under inadequate train samples
Chaos theory	Chaos theory	Rich data	Simplify water quality system that involving various influencing factors	Difficult to operate

Mathematical statistical method predicted the trends of changes of water quality by pre-existing water quality information, including single factor water quality prediction method (single index of affecting water quality prediction) and multi-factor comprehensive water quality prediction method (multi-index of affecting water quality prediction). The prediction result of multi-factor comprehensive water quality prediction method is better due to building multi-index water quality relation, compared to single factor water quality prediction method. It usually adopted multiple linear regression (Inoul, 1986；Aron, 1978), stepwise cluster analysis, stepwise regression analysis methods and so on. On the whole, prediction effect of regression analysis is better, but it still has some drawbacks: huge CPU consumption, equally treating data of different time, just emphasis on historical monitoring data fitting instead of extrapolation and so on. However, difficulty and accuracy of water quality prediction should be faced with due to many involved factors and a large amount of information needed for the method construction and difficulties on modeling (Li, 2006).

The grey system theory method bases on analysis and process of no obviously regular pre-existing water quality observed data to generate regular water quality monitoring data. And then fits generated data by differential equation and constructs grey system dynamic prediction model to perform long and medium term prediction of water quality. This method is suitable for the circumstances of less observed data because the request is not high for the historical observed data of water quality. This method is a trend prediction extrapolation method of single factor.

Artificial Neural Networks (ANN) is composed of a simple artificial neural connection into a complex network model to simulate human brain nerve network. It's high-dimensional property, self-organization, self-learning and strong adaptive ability and other ability to predict and early warn complex water quality problems. However, its disadvantages are obvious. On one hand, construction process of neural networks model is complex. On the other, the training of model requires a large amount of sample data of water quality monitoring. When it's applied to practical water environment management, this model hardly establishes water environment management decision support system due to its higher professional requirements for operators.

Chaology comes from ancient China. In 1963, Edward Norton Lorenz, an American meteorologist, proposed famous "butterfly effect" (Lorenz, 1963). Studies have shown that some completely certain nonlinear dynamic system, and even simple system described in a deterministic equations, in particular condition, may show the very complex, seemingly irregular random motion. Phase space reconstruction is application foundation of chaos theory. In terms of chaos theory and phase space reconstruction, complex features of river water quality is generally gasped. Complex multivariable system with multiple coupling relationship is converted into simple single-variable system. Short-term development trend of river water quality system is predicted by studying the inner evolution law of complex systems. Firstly, this method hypothesizes that the water quality monitoring data of time sequence is unlimited, and does not consider the effect of data noise. But in the process of actual water quality simulation analysis, due to the limitation of objective conditions, time sequence meeting above condition is hardly got and it's difficult to implement. So under the condition of incomplete actual monitoring data, how to choose the best length of time series and consider reducing the negative effect of noise remain to be further studied.

2.1.4 Comparison among mechanism models for water quality

Characteristics of mechanism water quality models are presented in Table 2-2.

Table 2-2. Characteristics of mechanism water quality models.

Model name	Model characteristics	Model dimensions	Whether source codes are open	Disadvantages
CE-QUAL-ICM	It can simulate water quality changes, water-sediment exchange process and spring "algal bloom" as well as summer oxygen deficit after the peak nutrient input.	One, two, three dimensions	Yes	Only simulate water quality
WASP	It simulates dynamics of surface water quality in which common and toxic pollutants migrate. It includes three independent simulation programs: hydrodynamic model, eutrophication model and toxic substance model, all of which can be operated independently. The model is widely applied.	One, two dimensions	Yes	No three-dimensional simulation
EFDC	It is the surface hydrodynamic model tool recommended by USEPA, including hydrodynamic module, water quality module, toxic substance module, sediment module, wind wave module and silt module. It is used to simulate one-dimensional, two-dimensional and three-dimensional flow fields of water system, substance transportation (including water temperature, salt content, viscosity and non-cohesive sands), ecological process and fresh inflow. It is a mature model and is widely used.	One, two, three dimensions	Yes	Non-commercial version has no visualization interface

MIKE	It includes hydrodynamic module, hydraulic structure operation module, rainfall runoff module, pollution load module, convective diffusion module, water quality module and GIS module. It is widely used.	One, two, three dimensions	No	Commercial license
DELFT3D	It includes six modules (water flow, hydrodynamic, wind wave, sediment, water quality and ecology). All of these modules are completely dynamic coupling. The model is extensively used.	Two, three dimensions	Partially	License limit
QUAL2k	It is applicable to chaotic dendritic river system and allows multiple intakes, drain outlets, tributary outlets and inlets. It is widely in river pollutant control and water quality management in Europe, North America and Asia.	Vertical one-dimension	Yes	Few simulation dimensions

CE-QUAL-ICM is a three-dimensional eutrophication model for the Chesapeake Bay and major tributaries developed by researchers at the U.S. Army Corps of Engineers Engineering Research and Development Center in Vicksburg, MS, USA. This model research was sponsored by the Chesapeake Bay Program Office, USEPA and the Army Engineer District, Baltimore of USA. The model computes constituent concentrations resulting from transport and transformations in well-mixed grids that can be arranged in one-, two-, or three-dimensional model. Thus, the model employs an unstructured cell system. The model computes and reports concentrations, mass transport, hydrodynamic transformations, and mass balances. Features to facilitate debugging include the ability to activate or deactivate model characteristics, diagnostic output, and volumetric and mass balances. Computations can be restarted following interruption due to computer failure or similar circumstances. CE-QUAL-ICM is coded in ANSI Standard FORTRAN F77. A multi-processor version is available but not generally released. The user must provide processors that prepare input files and process output for visualization. The model itself does not include the hydrodynamic model. Flows, volumes and diffusion coefficients must be specified externally and fed into the model. For a few steps of configurations, flows can be entered through an ASCII input file. For more advanced applications, hydrodynamic information is usually obtained from a hydrodynamic model such as the CH3D-WES model. The unstructured, finite volume structure of the model was chosen to facilitate linkage to a variety of hydrodynamic models (Cole et al., 1995).

Water Quality Analysis Simulation Program (WASP) is a generalized framework for modeling contaminant fate and transport in surface waters. Based on the flexible compartment modeling approach, it can be applied in one, two or three dimensions and is designed to permit easy substitution of user-written routines into program structure. Problems studied using WASP framework include BOD and DO dynamics nutrients and eutrophication, bacterial contamination, and organic chemical and heavy metal contamination. Two WASP models are provided: Toxics, TOXI, combines kinetic structure with WASP transport structure and simple sediment balance algorithms to predict dissolved and sorbed chemical concentrations in the bed and overlying waters; dissolved oxygen/eutrophication, EUTRO, combines kinetic structure with WASP transport structure to predict DO and phytoplankton dynamics affected by nutrients and organic material. The model limitations are: Requires external hydrodynamic model to provide flow file for solving advection; The file size might be very large in several gigabytes for long-term simulation; User specified dispersion coefficient and temperature; First-order UPWIND difference in space may cause significant numerical diffusion; Over-simplified sediment flux calculation; No periphyton or macroalgae; Sediment transport processes are not related to shear stress. A significant amount of WASP applications can be found in technical reports, journal and conference papers (Ambrose et al., 1993).

MIKE is an implicit finite difference model for one dimensional unsteady flow computation and can be applied to looped networks and quasi-two dimensional flow simulation on floodplains. The model has been designed to perform detailed modelling of rivers, including special treatment of floodplains, road overtopping, culverts, gate openings and weirs. MIKE 11 is capable of using kinematic, diffusive or fully dynamic, vertically integrated mass and momentum equations (the Saint Venant equations). The solution of the continuity and momentum equations is built on an implicit finite difference scheme. This scheme is structured so as to be independent of the wave description specified (i.e. Kinematic, Diffusive or dynamic). Boundary types include water level (h), Discharge (Q), Q/h relation, wind field, dam break, and resistance factor. The water level boundary must be applied to either the upstream or downstream boundary condition in the model. The discharge boundary can be applied to either the upstream or downstream boundary condition, and can also be applied to the side tributary flow (lateral inflow). The lateral inflow is used to describe runoff. The Q/h relation boundary can only be applied to the downstream boundary. MIKE 11 is a modelling package for the simulation of surface runoff, flow, sediment transport, and water quality in rivers, channels, estuaries, and floodplains. The most commonly applied hydrodynamic (HD) model is a flood management tool simulating the unsteady flows in branched and looped river networks and quasi two-dimensional flows in floodplains. When using a fully dynamic wave description, MIKE 11 HD solves the equations of conservation of continuity and momentum (the Saint Venant equations). The solutions to the equations are based on the following assumptions: The water is incompressible and homogeneous (i.e. negligible variation in density) ; The bottom slope is small, thus the cosine of the angle it makes with the horizontal may be taken as 1; The wave lengths are large compared to the water depth, assuming that the flow everywhere can be assumed to flow parallel to the bottom (i.e. vertical accelerations can be ignored, and a hydrostatic pressure variation in the vertical direction can be assumed); The flow is sub-critical (a super-critical flow is modelled in MIKE 11; however, more restrictive conditions are applied) (Danish Hydraulics Institute, 1993; Kamel, 2008).

Delft3D has been developed by Deltares as a unique, fully integrated computer software suite for a multi-disciplinary approach and 3D computations for coastal, river and estuarine areas. It has the ability of simulating flows, sediment transports, waves, water quality, morphological developments and ecology. The Delft3D suite is composed of several modules, grouped around a mutual interface, while being competent to interact with one another. An 'Online' approach was adopted in Delft3D to incorporate integrated flow, sediment transport and bathymetry updates are executed at each time step (Roelvink, 2006). Each of these modules use the others' output as input of next time step. It is a multi-dimensional (2D or 3D) hydrodynamic simulation program that calculates non-steady flow and transport phenomena in shallow water. Shallow-water means shallow seas, coastal areas, estuaries, lagoons, rivers and lakes, where horizontal spatial and temporal scales are much larger than vertical scales (Lesser et al., 2004). The flow and transport phenomena result from tidal and meteorological forcing on a rectilinear or a curvilinear, boundary fitted grid. The 2D-schematization averages a body of water in depth, whereas in 3D simulations, the hydrodynamic module applies the so-called 'sigma co-ordinate transformation' on the vertical axis, which results in a smooth representation of the bottom topography. (Deltares, 2013) In the staggered grid, which was applied in Delft3D-FLOW, not all the properties of the grid cell, such as water level, water depth, the velocity components or concentration of sediments, were defined at the same location in the numerical grid. Delft3D solves shallow water equations. The system of equations consists of horizontal momentum equations, the continuity equation, the transport equation, and a turbulence closure model. Delft3D offers various transport formulations by default, while also enabling users to define their own and call them from Delft3D-FLOW. These transport formulations include Van Rijn, Engelund-Hansen, and Ackers-White etc. Each is an empirical relation designed for different situations, such as wave conditions and sediment loads (Chen, 2014).

QUAL2K is an one-dimensional river and stream water quality model that is an upgraded version of the QUAL2E model (Brown et al., 1987). The

QUAL2K framework, which was promoted by the US Environmental Protection Agency, can simulate the migration and transformation of conventional pollutants. The model considers the stream as an one-dimensional channel with steady flow that is non-uniform and considers the influence of point source and nonpoint source pollution loads. QUAL2K can simulate the migration and transformation of a wide variety of constituents including dissolved oxygen, temperature, biochemical oxygen demand, organic nitrogen, NH_3N, nitrate nitrogen, total nitrogen, sediment oxygen demand, organic phosphorus, inorganic phosphorus, total phosphorus, phytoplankton and algae. The illustrations and uses of this model are detailed in the QUAL2K user's manual (Chapra et al., 2008). The model can also simulate some other factors, including pH, alkalinity and pathogenic bacteria. (Zhang et al., 2012).

2.2 Introduction and fundamental theories of EFDC

2.2.1.Introduction to EFDC

The Environmental Fluid Dynamics Code (EFDC) is the comprehensive water quality mathematical model software developed by John Hamrick et al. (Virginia Marine Scientific Institute, Marine School of College of William and Mary) based on subsidizes of EPA (the United States) and integration of many mathematical models (Hamrick, 1991).

The technical report of EFDC software was released and applied to the research project of flow field near the Newport News Port, Virginia, the United States in 1992 (Hamrick, 1992). In 1993, it was used to study early pollutant mixing and dilution in York River, the United States (Hamrick, 1993).In 1994, it was used in the research project of water conservation district in South Florida (Hamrick, 1994). EFDC software has been being perfected since Hamrick joined into Tetra Tech Company in 1996. External supports to development, maintenance and application of EFDC software mainly come from associated research departments of EPA. Hamrick published the instruction manual of

EFDC software in 1996 (Hamrick, 1996). The Benchmark test report was released on the next generation of environmental model and computation model seminar which was held in the National Environmental Super computation Center of the United States in 1997 (Wu et al., 1997). In 2002, the model assessment center of EPA launched the first open edition of EFDC software and provided download of executable programs and related technical files. According to the disclosure statement, EFDC software is one of the most advanced hydrodynamic models that can be used for one-dimensional, two-dimensional and three-dimensional simulation of water ecosystem.

Moreover, the model assessment center of EPA released the improved edition of EFDC software. At present, EPA is developing pre-processing software of EFDC, which has two major functions: modeling interface and curve meshing. In addition, DIS Corporation and Aquaveo Company (USA) improved EFDC model based on source codes and developed pre-processing as well as post-processing commercial software (Jiang Peng et al., 2014). Through continuous R&D, maintenance, improvement and application, EFDC software has been used in regional environmental evaluation and management successfully. It is the technical support to monitoring and involves various types of water bodies, including river, lake, wetland, estuary and sea.

2.2.2.State variables and components of EFDC

EFDC can process a large suite of model state variables (Table 2-3). The nitrate state variable in the model represents the sum of nitrate and nitrite nitrogen. The three variables (salinity, water temperature, and total suspended solids) needed for computation of the state variables are provided by EFDC hydrodynamic model. The interactions among the state variables is illustrated in Figure 2-1. The kinetic processes included in EFDC water quality model are mostly from the Chesapeake Bay three-dimensional water quality model, CE-QUAL-ICM (Cerco et al., 1995).

Table 2-3. EFDC model water quality state variables.

1) cyanobacteria	12) labile particulate organic nitrogen
2) diatom algae	13) dissolved organic nitrogen
3) green algae	14) NH_3N
4) refractory particulate organic carbon	15) nitrate nitrogen
5) labile particulate organic carbon	16) particulate biogenic silica
6) dissolved organic carbon	17) dissolved available silica
7) refractory particulate organic phosphorus	18) chemical oxygen demand
8) labile particulate organic phosphorus	19) dissolved oxygen
9) dissolved organic phosphorus	20) total active metal
10) total phosphate	21) fecal coliform bacteria
11) refractory particulate organic nitrogen	22) macroalgae

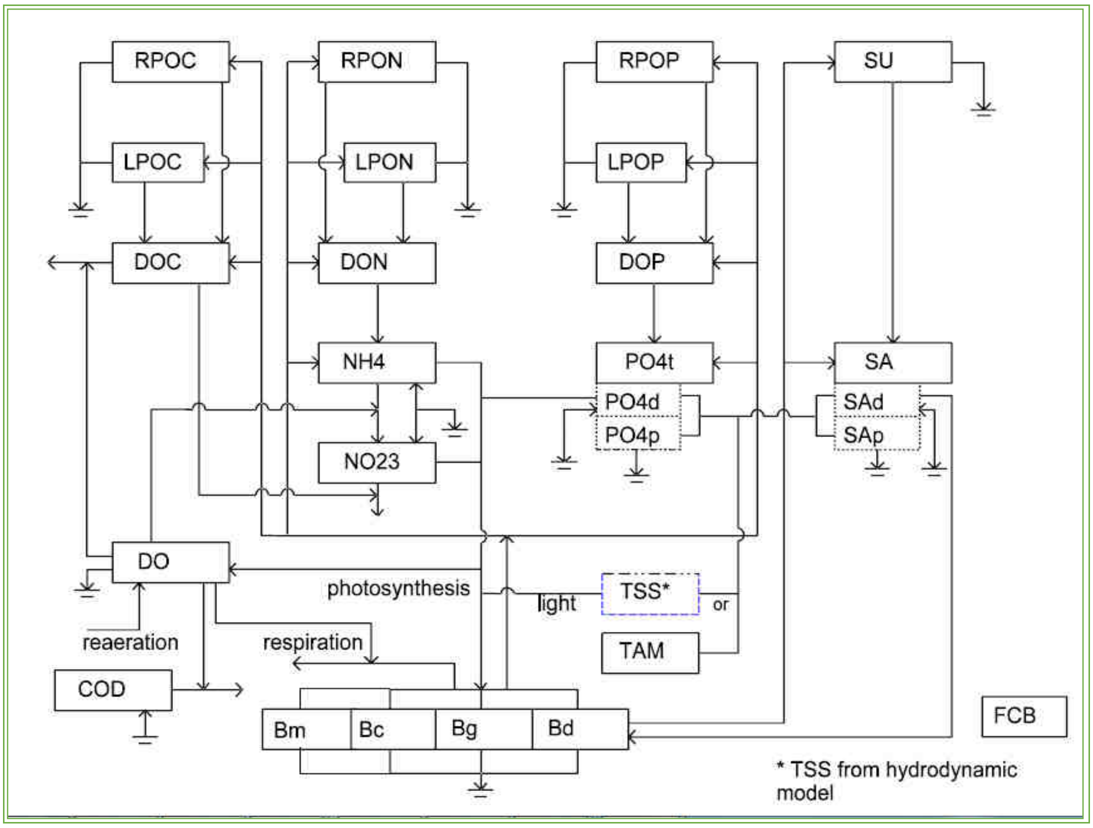


Figure 2-1. Schematic diagram for EFDC water quality model.

The structure of the EFDC includes four major modules: (1) a hydrodynamic model, (2) a water quality model, (3) a sediment transport model, and (4) a toxics model (figure 2-2). The EFDC hydrodynamic model itself, which was used for this study, is composed of six transport modules including dynamics, dye, temperature, salinity, near field plume, and drifter (figure 2-3). Various products of the dynamics module (i.e., water depth, velocity, and mixing) are directly coupled to the water quality, sediment transport, and toxics models as shown in the following figures. Schematic diagrams for the water quality model, the sediment transport model, and the toxics model are shown in figures 2-4, 2-5, and 2-6, respectively (Park et al., 1995).

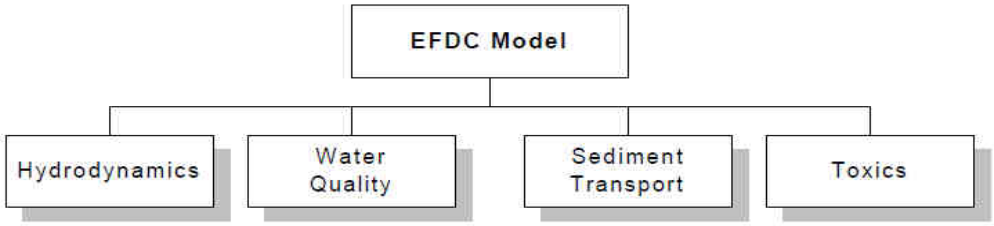


Figure 2-2. Primary modules of the EFDC model.

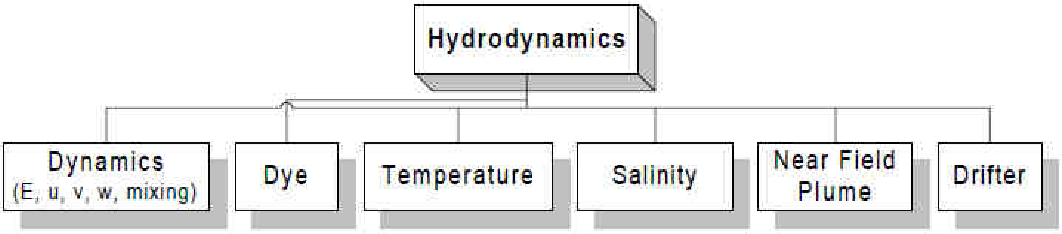


Figure 2-3. Structure of the EFDC hydrodynamic model.

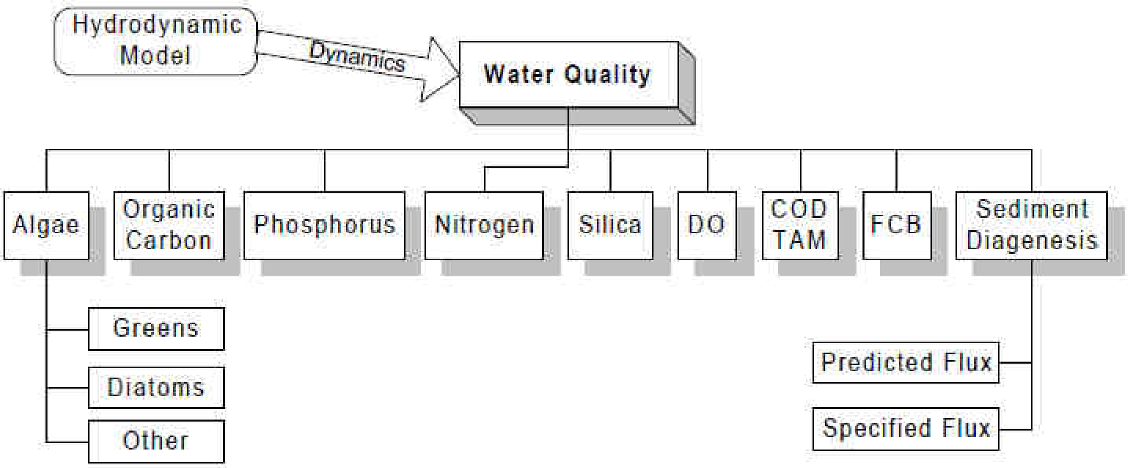


Figure 2-4. Structure of the EFDC water quality model.

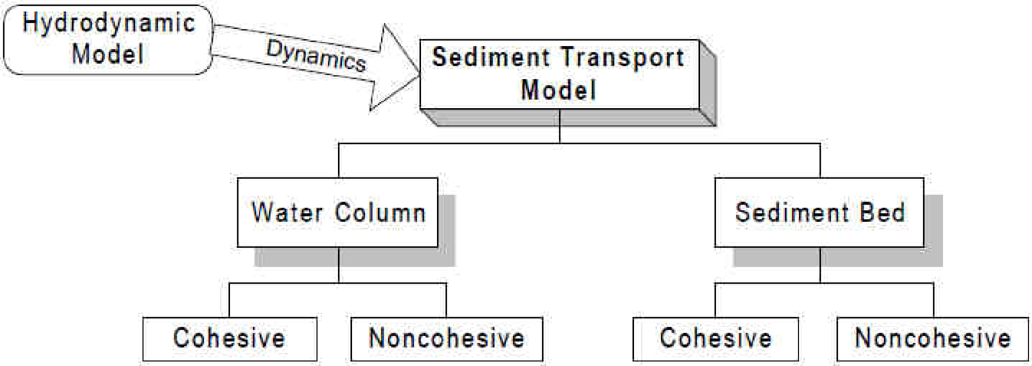


Figure 2-5. Structure of the EFDC sediment transport model.

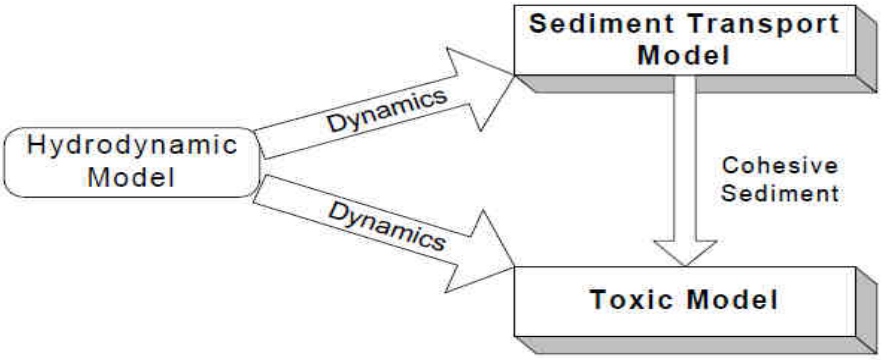


Figure 2-6. Structure of the EFDC toxics model.

2.2.3.Fundamental equations in EFDC

The hydrodynamic equations in EFDC are based on a three-dimensional incompressible, graded-density turbulence boundary layer equation set, including the momentum equation, continuity equation, and material transport equation. The Boussinesq hypothesis (Boussinesq, 1903) is often adopted to facilitate the processing of buoyancy lift terms caused by density contrast. The governing equations are given after the two types of transformation, as follows.

Momentum equations:

$$\partial_t(mHu)+\partial_x(m_yHuu)+\partial_y(m_xHvu)+\partial_z(mwu)-(mf+v\partial_x m_y-u\partial_y m_x)Hv$$
$$=-m_yH\partial_x(g\zeta+p)-m_y(\partial_x h-z\partial_x H)\partial_z p+\partial_z(mH^{-1}A_V\partial_z u)+Q_u \quad (2\text{-}1)$$

$$\partial_t(mHv)+\partial_x(m_yHuv)+\partial_y(m_xHvv)+\partial_z(mwv)+(mf+v\partial_x m_y-u\partial_y m_x)Hu$$
$$=-m_xH\partial_y(g\zeta+p)-m_x(\partial_y h-z\partial_y H)\partial_z p+\partial_z(mH^{-1}A_V\partial_z v)+Q_v \quad (2\text{-}2)$$

$$\partial_z p = -gH(\rho - \rho_0)\rho_0^{-1} = -gHb \tag{2-3}$$

Continuity equations:

$$\partial_t(m\zeta) + \partial_x(m_y Hu) + \partial_y(m_x Hv) + \partial_z(mw) = 0 \tag{2-4}$$

$$\partial_t(m\zeta) + \partial_x(m_y H\int_0^1 udz) + \partial_y(m_x H\int_0^1 vdz) = 0 \tag{2-5}$$

$$\rho = \rho(p, S, T) \tag{2-6}$$

Transport equations:

$$\partial_t(mHC) + \partial_x(m_y HuC) + \partial_y(m_x HvC) + \partial_z(mwC) = \partial_z(mH^{-1}A_b\partial_z C) + Q_C \tag{2-7}$$

$$\partial_t(mHT) + \partial_x(m_y HuT) + \partial_y(m_x HvT) + \partial_z(mwT) = \partial_z(mH^{-1}A_b\partial_z T) + Q_T \tag{2-8}$$

In these equations, u and v are the horizontal velocity components in the curvilinear, orthogonal coordinates x and y, m_x and m_y are the square roots of the diagonal components of the metric tensor, $m = m_x m_y$ is the Jacobian or square root of the metric tensor determinant, A_v is the vertical turbulent viscosity, A_b is the vertical turbulent diffusivity, f is the Coriolis parameter, p is the physical pressure, The density, r, is in general a function of temperature, T, and salinity or water vapor, C, in hydrospheric and atmospheric flows respectively and can be a weak function of pressure, consistent with the incompressible continuity equation under the anelastic approximation, ρ_0 is the reference density, Q_u and Q_v are the momentum source-sink terms, the source and sink terms, Q_c and Q_t include subgrid scale horizontal diffusion and thermal sources and sinks. The vertical velocity, with physical units, in the stretched, dimensionless vertical coordinate z is w, and is related to the physical vertical velocity w^* by:

$$w = w^* - z(\partial_t\zeta + um_x^{-1}\partial_x\zeta + vm_y^{-1}\partial_y\zeta) + (1-z)(um_x^{-1}\partial_x h + vm_y^{-1}\partial_y h) \tag{2-9}$$

The system of eight equations (equations (2-1)-(2-8)) provides a closed system for the variables u , v , w, p, z, r, C, and T, provided that A_v and A_b diffusivity and the source and sink terms are specified. The total depth, $H =$

$h + z$, is the sum of the depth below and the free surface displacement relative to the undisturbed physical vertical coordinate origin. The continuity equation (2-4) has been integrated with respect to z over the period $(0,1)$ to produce the depth integrated continuity equation (2-5) using the vertical boundary conditions, $w = 0$ at $z = (0,1)$, which follows from the kinematic conditions and equation (2-9).

2.2.4.Turbulence closure model

To obtain A_v and A_b, the second moment turbulence closure model will be used, the model relates A_v and A_b to the turbulent intensity, q a turbulent length scale, l and a Richardson number R_q by:

$$A_v = \phi_v ql = 0.4(1+36R_q)^{-1}(1+6R_q)^{-1}(1+8R_q)ql \quad (2\text{-}10)$$

$$A_b = \phi_b ql = 0.5(1+36R_q)^{-1} ql \quad (2\text{-}11)$$

$$R_q = \frac{gH\partial_z b}{q^2}\frac{l^2}{H^2} \quad (2\text{-}12)$$

Where the so-called stability functions ϕ_v and ϕ_b account for reduced and enhanced vertical mixing or transport in stable and unstable vertically density stratified environments, respectively. The turbulence intensity and the turbulence length scale are determined by a pair of transport equations:

$$\frac{\partial(mHq^2)}{\partial t}+\frac{\partial(m_yHuq^2)}{\partial x}+\frac{\partial(m_xHvq^2)}{\partial y}+\frac{\partial(mwq^2)}{\partial z}=\frac{\partial}{\partial z}(m\frac{1}{H}A_q\frac{\partial q^2}{\partial z})+Q_q \quad (2\text{-}13)$$
$$+2m\frac{1}{H}A_v\left(\left(\frac{\partial u}{\partial z}\right)^2+\left(\frac{\partial v}{\partial z}\right)^2\right)+2mgA_b\frac{\partial b}{\partial z}-2mH\frac{1}{B_1 l}q^3$$

$$\frac{\partial(mHq^2l)}{\partial t}+\frac{\partial(m_yHuq^2l)}{\partial x}+\frac{\partial(m_xHvq^2l)}{\partial y}+\frac{\partial(mwq^2l)}{\partial z}=\frac{\partial}{\partial z}(m\frac{1}{H}A_q\frac{\partial q^2l}{\partial z})+Q_l \quad (2\text{-}14)$$
$$+m\frac{1}{H}E_1A_v\left(\left(\frac{\partial u}{\partial z}\right)^2+\left(\frac{\partial v}{\partial z}\right)^2\right)+mgE_1E_3lA_b\frac{\partial b}{\partial z}-mH\frac{1}{B_1}q^3\left(1+E_2\frac{1}{(KL)^2}l^2\right)$$

$$\frac{1}{L}=\frac{1}{H}\left(\frac{1}{z}+\frac{1}{(1-z)}\right) \quad (2\text{-}15)$$

WhereB_1, E_1, E_2, and E_3 are empirical constants, Q_q and Q_l are additional source-sink term such as subgrid scale horizontal diffusion. The vertical diffusivity, A_q is in general taken equal to A_v (Harmick, 1992).

2.2.5.Numerical solution techniques

In the above-mentioned equation set, second-order accuracy finite difference is adopted in the solution of equations (2-1), (2-2), and (2-4). Staggered grid scatter is adopted horizontally (Mellor et al., 1982). Second-order accuracy finite difference in three time level scheme is adopted in time integration. The solution is divided into the internal mode and the external mode, i.e. free surface gravity waves and shear stresses are solved in splitting methods (Arakawa et al., 1997). Semi-implicit difference schemes are adopted in the solution in the external mode. Two-dimensional water level elevation is calculated simultaneously. In this mode, pre-processing is conducted in the conjugate gradient method before solution (Madala et al., 1977). The solution method allows large-scale time step calculation. Time step is only constrained by the explicit central difference stability criterion or the high-order windward advection algorithm of the non-linear accelerating algorithm (Sheng et al., 1982). Implicit difference scheme with consideration of vertical diffusion is adopted in the solution to the internal model (Smolarkiewicz et al., 1993).

The three time level and step by step algorithm resolve the solution (Sheng, 1986) of temperature, q^2 and $q^2 l$ in the material transport equation. To minimize numerical diffusion, the multidimensional positive definite advection transport algorithm is adopted in the model (Smolarkiewicz et al., 1986; Smolarkiewicz et al., 1990). The algorithm adopts first-order accuracy spatially and second-order accuracy temporally.

2.3 Software foundation of establishing the water quality model for Mudan River

Although abundant fruits concerning water quality warning technology have been achieved, both mechanism and non-mechanism models still have some practical application limitations. Non-mechanism water quality models are widely used for their simple modeling and strong adaptation to boundary conditions. However, they depend highly on data. Establishment, verification and application of models are closely related with data characteristics. Non-mechanism models mainly focus on a certain object and are established based on specific data, resulting in the poor extensibility (Song, 2013). Mechanism water quality models take physical, chemical and biological factors that will influence water quality into account and could describe the mechanism of pollutant diffusion clearly and predict water quality accurately. Nevertheless, they require enough data during actual modeling, such as river bed elevation data, hydrologic data, meteorological data, etc. Lacking basic data of most rivers in China for a long time in the past, further researches on mechanism models involving relatively complete data are needed.

It can conclude from above analysis that existing water quality early warning systems have both advantages and disadvantages in modeling and applications. There are relatively adequate data about the Mudan River. Its water quality SEWS was established by a mechanism model. EFDC was used to establish the water quality simulation model for the Mudan River, which was integrated into the GIS environment in view of its following advantages: 1) EFDC is the surface water simulation tool that is strongly recommended by EPA of the United States. Its accuracy in hydrodynamic simulation (Jiang et al., 2009) and water quality simulation is accepted in the academic circles to a certain extent. 2) Document input is easy to understand and can make better coupling of hydrodynamic, water quality and sediment modules. Transformation between ports of different models can be neglected. 3) As a completely open source software, EFDC is not only good for integration of the model and GIS functional components as well as further development (Huang et al., 2015), but also conducive to developing a distributed SEWS. To simplify the intensive model computation, improve extensibility of model simulation and reduce hardware resource waste, distributed dynamic elastic computation of water quality model was studied when establishing the SEWS.

2.4 Chapter conclusion

This chapter analyzed development phase, definition and classification of water quality model, and introduced the common technology method of water quality simulation and early warning, including mechanism model and non-mechanism model. Non-mechanism model mainly adopted mathematical statistics method, grey system theory, neural network method and so on. However, mechanism model mainly contained WASP, EFDC, MIKE, DELFT3D, QUAL2k and so forth. This chapter compared the advantages and disadvantages among water quality simulation methods of mechanism models and non-mechanism models, and considered features of research objective and advantages of EFDC, then selected EFDC as the basis to establish the water quality model in Mudan River by the specific research requirements of objective and distributed system development. Finally, this chapter introduced the model concept and theoretical foundation involved by EFDC.

Chapter 3

Building and verification of water quality model for Mudan River

3.1 Overview of studies area

3.1.1 Hydro-meteorological conditions

Mudan River is located in the sub-Songnen Plain of northeast China climate, having continental climate of cold-temperate zone with obvious seasonal changes. According to the observations from Mudan River Meteorology Station (N44.57°, E129.6°), changes of air temperature and precipitation in Mudan City is relatively consistent. Namely, temperature in July and August is higher, with highest mean temperature over 21℃. At the same period, monthly maximum mean precipitation has reached 123.95mm. However, air temperature in January is lower, the minimum mean air temperature is -17.82℃. At the same period, rainfall is at its lowest with monthly mean precipitation being 5.45mm, as shown in Figure 3-1.

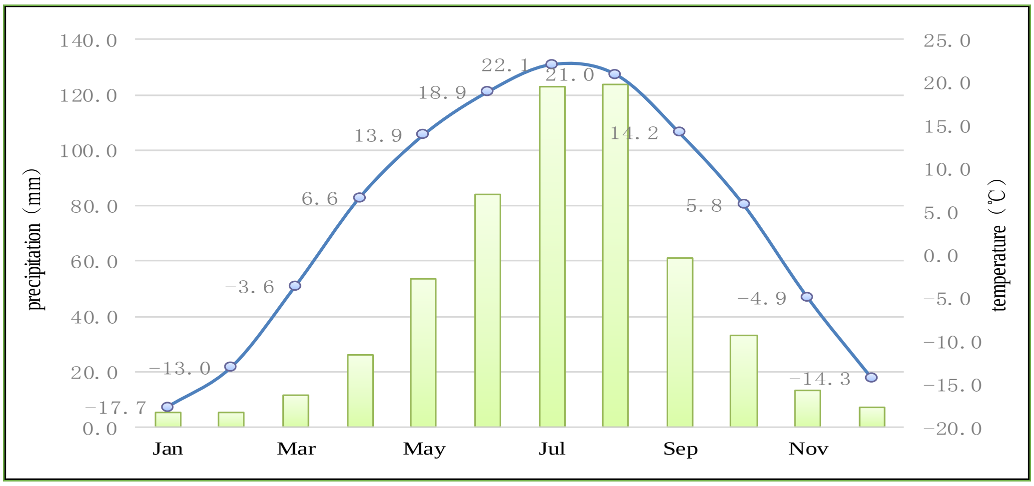


Figure 3-1. Statistical curve of monthly mean temperature and precipitation of Mudan City (1951-2010).

(Data source: China Meteorologic Data Website)

Figure 3-1 shows that meteorologic conditions of Mudan city are characteristic of obvious seasonal features. Therefore, water transportation and diffusion rules in the urban area of Mudan River have been studied in terms of water periods. According to statistical analysis, high flow period is from July to September, middle flow period includes May, June and October while December to April of next year is low flow period which is also the ice-covered period.

3.1.2 Hydrology features of trunk stream

Mudan River is the second largest trunk stream of Songhua River and originated from Mudanling of Jilin Changbai Mountain. The river flows from south to north, with a total length of 726km, river width of 100-300m, water depth of 1.0-5.0m, total drop of 1007m and mean slope drop of 1.39‰. Mudan River watershed flows through Heilongjiang Province and Jilin Province, with a total area of 37,023km^2, of which, Heilongjiang Province takes an area of 28,543km^2,77% of the total area. The Mudan River streams from south to north crossing Dunhua city of Jilin Province, Ning'an, Hailin, Mudan, Linkou and Yilan among others of Heilongjiang Province, flowing into Songhua River at the south of Yilan County. According to statistical data in many years' observation, mean flow of Mudan River estuary is 258.5m^3/s, mean runoff is 5.26 billion m^3, the maximum runoff is 14.9 billion m^3, being approximately 10% of total runoff of Songhua River water(Wang et al., 2013).

Tributaries of Mudan River are distributed relatively evenly along the two banks, with a treelike water system (Figure 3-2). There are a lot of torrential tributaries in high flow period. Downward from Mudanjiang City, left tributaries are often perpendicular to trunk stream. The largest tributary is Hailang River with a watershed being 1/7 of total watershed in area. The mean runoff depth of Mudan River trunk stream does not vary too much. Upper stream is larger than lower stream. The runoff depth is from 263mm to 214mm. Runoff depth of left tributaries is larger than that of right ones. The mean runoff depth of upper stream of Hailang River is 317-391mm, coefficient of annual runoff depth variation is 0.35-0.4. Mean runoff of Mudanjiang Station is 5.26×108m^3.

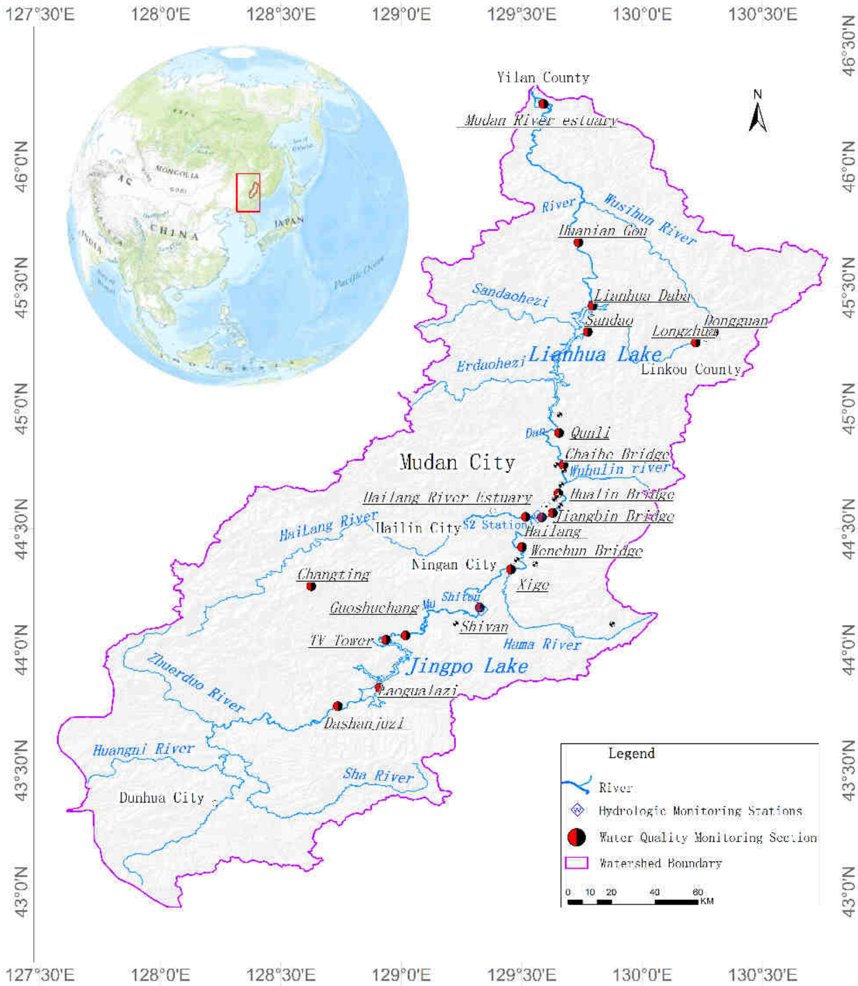


Figure 3-2. Diagram of water quality monitoring sections for Mudan River.

According to the analysis of hydrologic data of 1999 to 2008 obtained by Mudanjiang S2 station and Shitou hydrologic monitoring station, Mudan River is a seasonal river, the flow in flood period from May to September is much larger than that in low flow period. The water level curve and flow change curve obtained by Mudanjiang S2 hydrologic monitoring station in 2008 is shown in Figure 3-3, which presents that the flow in low flow period is between 30-50m^3/s, while that in flood period is often higher than 200m^3/s, even higher than 400m^3/s in a lot of cases. Water level variation is between 224-226mm, the trend of which is consistent with that of water flow. Larger flow corresponds to high water level and vice versa. In 2008, the maximum flood flow was 515m^3/s on August

13, corresponding the highest water level of 225.8m. The lowest flow was 23.6m^3/s on February 8, corresponding the water level of 224.59m.

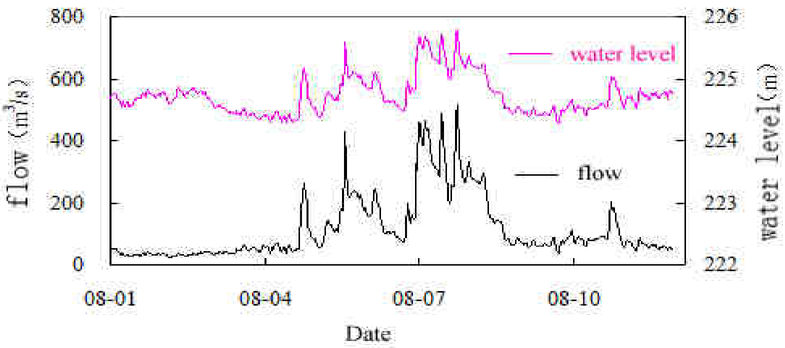


Figure 3-3. Variation of water level and flow in 2008 observed by Mudanjiang S2 hydrologic monitoring station.

Flow-water level relation curve obtained by Mudanjiang S2 hydrologic monitoring station in 2008 is shown in Figure 3-4. It can be found that flow increases with the increase of water level. In middle and large flows, the two parameters present a good correlation, majority of the points are on a smooth curve, which is similar to the rules of most rivers. It should be noted that in low flow condition, a lot of points deviate from the curve, indicating that a different water level-flow relation exists in such circumstance. Low flow period of Mudan River is around winter and spring with rather low temperature and frozen water, which can increase the water resistance. It can also be demonstrated in Figure 3-4 that all the deviated points located to the lower right of the curve, suggesting that flow capacity is decreased under given water level. It means that a larger flow area is required for the same flow. Because these points are rather scattered, it is difficult to find a specific water level-flow relation. Therefore, under ice-covered conditions, the water features of Mudan River is quite complicated.

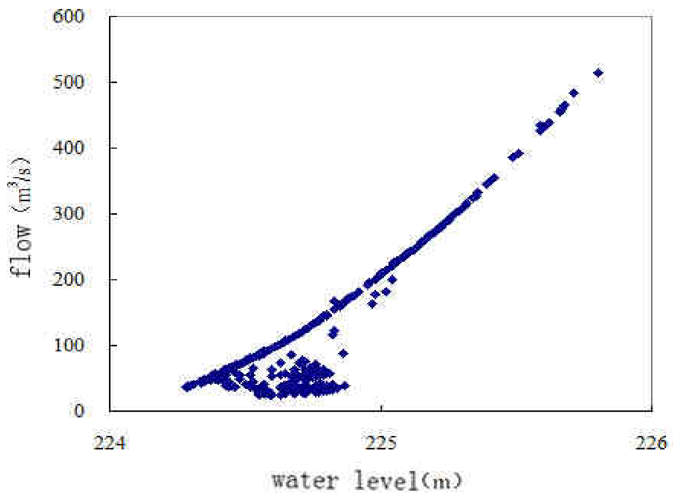


Figure 3-4. Relation curve of flow-water level obtained by Mudanjiang S2 Hydrologic Monitoring Station in 2008.

3.2 Current water pollution analysis of Mudan River

3.2.1 Water function zones and water quality monitoring section deployment

According to water function zones of Mudan River, fifteen regular water quality monitoring sections have been laid along the trunk stream of Mudan River watershed. Among the fifteen sections, five are nation-controlled, four are province-controlled, five are city-controlled and one is spared. Two sections Changting River and Hailang River estuary have been laid on the key tributary, Hailang River. Two sections of Dongguan and Longzhua have been laid on Wusihun River. The position information of each monitoring section is illustrated in Figure 3-2.

3.2.2 Analysis of polluted water drainage features

(1) Distribution of sewage outlets along Mudan River

There are thirteen major waste outlets along the trunk stream of Mudan River. Four are located in Ning'an City section, five in Mudanjiang City section, three in Hailin City and one in Linkou County. The spatial distribution of these outlets are shown in Figure 3-5.

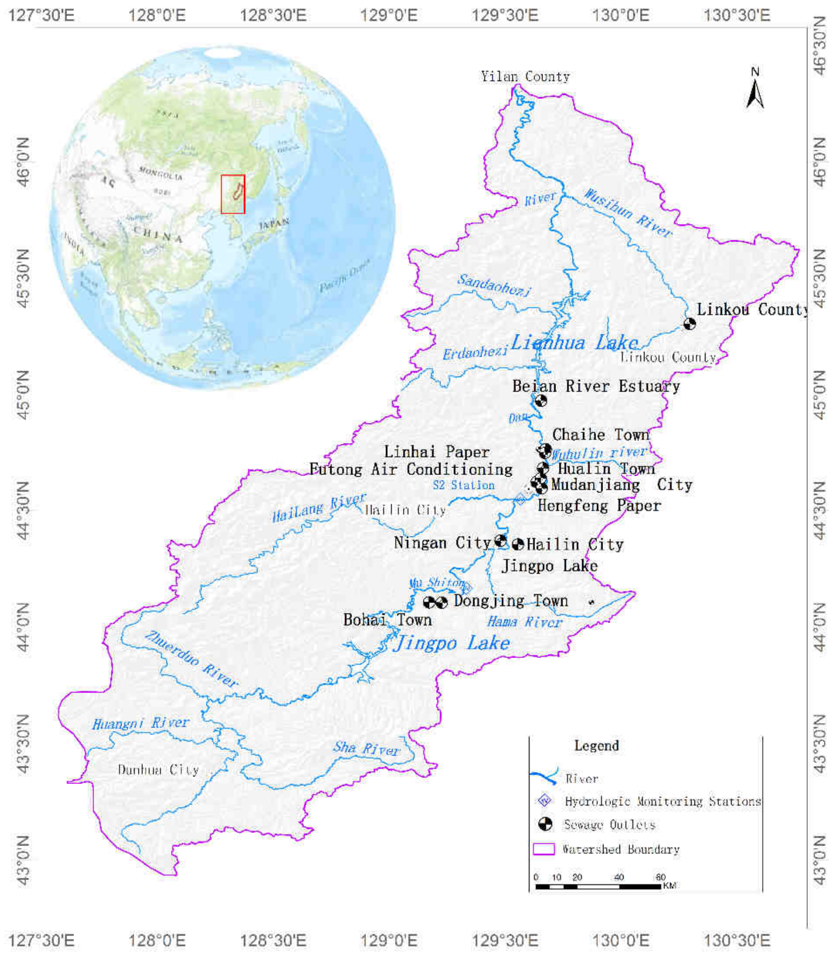


Figure 3-5. Sewage Outlets Distribution along Mudan River.

(2) Analysis of current sewage discharge

On the basis of pollution discharge data (COD and NH_3N) monitored from all the sewage outlets in 2014, statistical analysis have been done in terms of pollutant concentration variation, contribution rate of waste water discharge and contribution rate of pollutant discharge, which are illustrated in Figure 3-6 and Figure 3-7. It can be found in the two figures that except for Beian River estuary, the maximum and minimum values of COD and NH_3N of which are to a large extent deviated from their respective mean values, the pollutant concentrations in other outlets do not vary too much. This is mainly because that the waste water sources of Beian River are various and complicated, leading to a varying concentration. However, the pollution sources of other outlets are relatively simple, leading to a stable pollutant concentration.

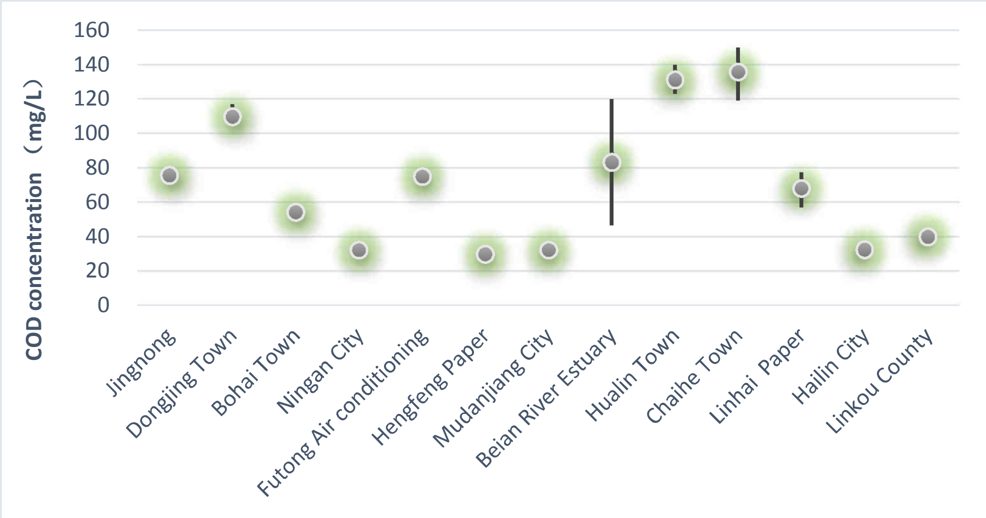


Figure 3-6. COD variation statistics for all the waste discharge outlets in 2014.

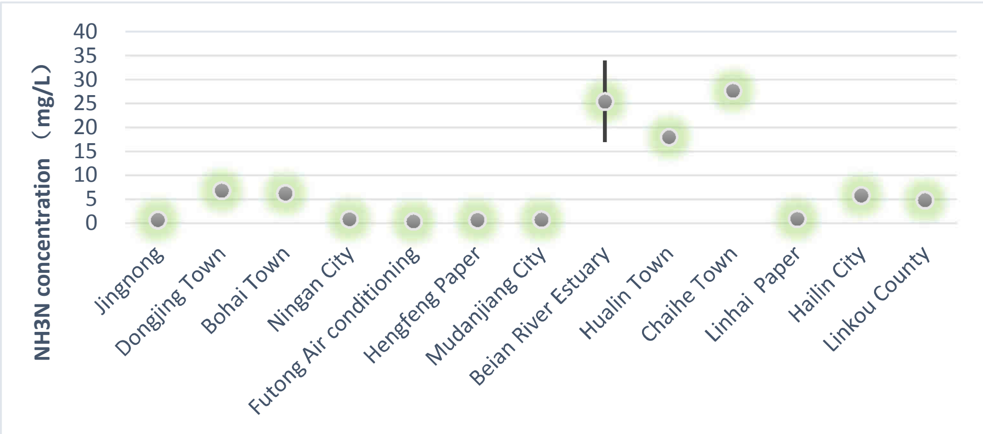


Figure 3-7. NH3N concentration variation statistics for all the sewage outlets in 2014.

1) Waste water discharge volume

It can be found in Figure 3-8 that sewage outlet of Mudanjiang City Waste Water Plant is the largest one in the watershed, the discharge volume of which takes up 58.81% of the total volume. Mudanjiang City Waste Water Plant mainly receives and processes sanitary water from urban area. The processed sanitary water discharge volume takes up 99% of the total volume discharged through this outlet. Some other large outlets include waste water discharge outlets of Ning’an Waste Water Plant, Hailin Waste Water Plant and Linkou County, taking up 10.07%, 8.74% and 9.46% of the total discharged volume, respectively. The waste water discharged by Hengfeng Paper, takes up 4.85% of the total volume, being regarded as one large outlet. However, the pollutant concentration is low and has reached relevant regulation. The waste water discharge volume from Beian River only takes up 2.39% of the total volume. Nevertheless, due to the high pollutant concentration, the contribution rate of pollutant discharge of Beian River is much higher than that of other ones. The rest outlets discharge only less than 1% of the total volume.

2) Sewage Discharge Volume

It can be found in Figure 3-9 that among these waste discharge outlets along Mudan River, Mudanjiang Waste Water Plant discharges the largest volume of COD, with a contribution rate of 51.17%, followed by Linkou County, Ning'an Waste Water Plant and Hailin Waste Water Plant, contribution rates of which are 10.22%,8.76% and 7.63%, respectively. The next ones are Beian River, Hengfeng Paper and Linhai Paper, contribution rates of which are 5.36%, 3.91% and 3.69%, respectively. The rest outlets only contribute a small portion.

It can be observed from Figure 3-10 that Beian River contributes the largest NH_3N discharge volume with a contribution rate of 24.25%. The next ones are Hailin Waste Water Plant, Mudan River Waste Water Plant and Linkou Country, whose contribution rates are 20.25%, 18.29% and 18.06% respectively. The next two are Chaihe Town and Ning'an Waste Water Plant, whose contribution rates are 6.42% and 3.33% respectively. The rest ones contribute a little.

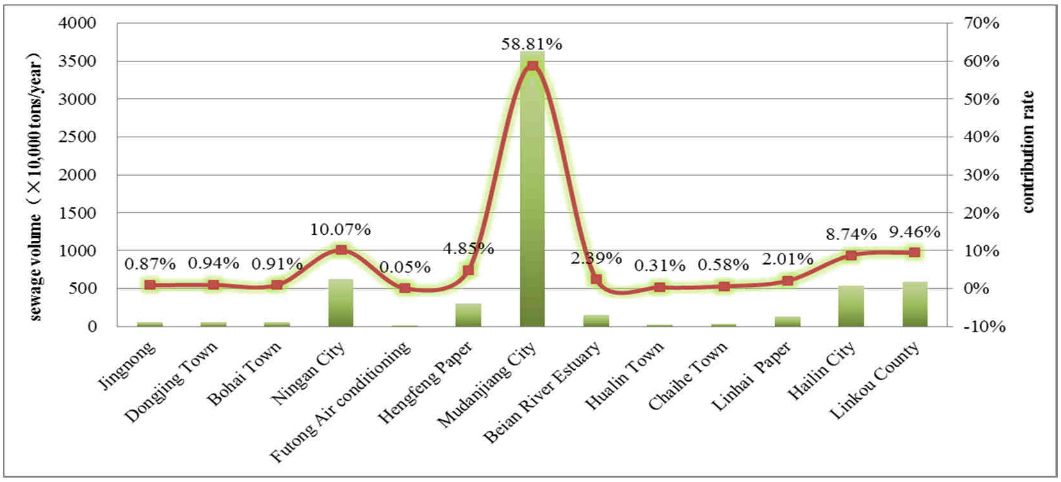


Figure 3-8. Waste water discharge volume and contribution rate of each discharge outlet along Mudan River.

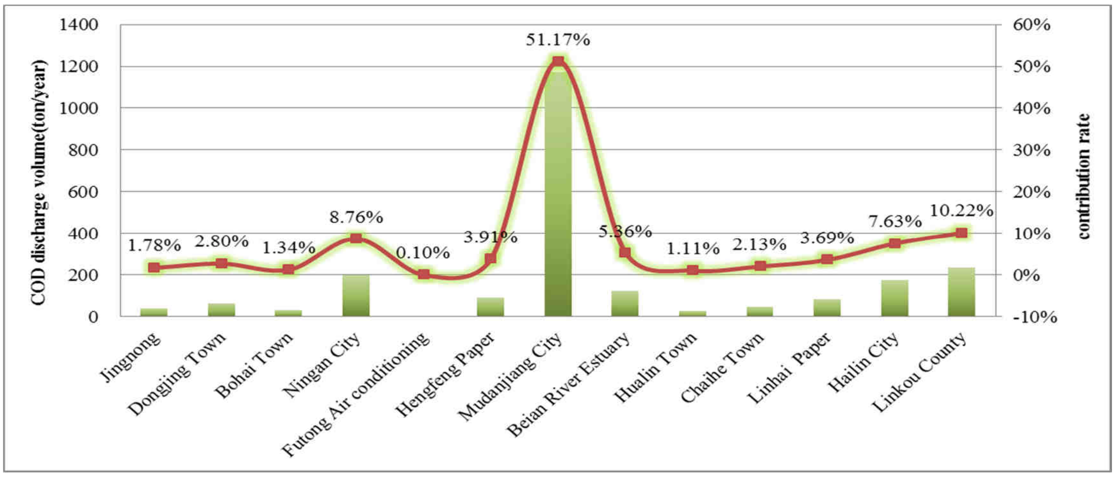


Figure 3-9. COD discharge volume and contribution rate of each discharge outlet along Mudan River.

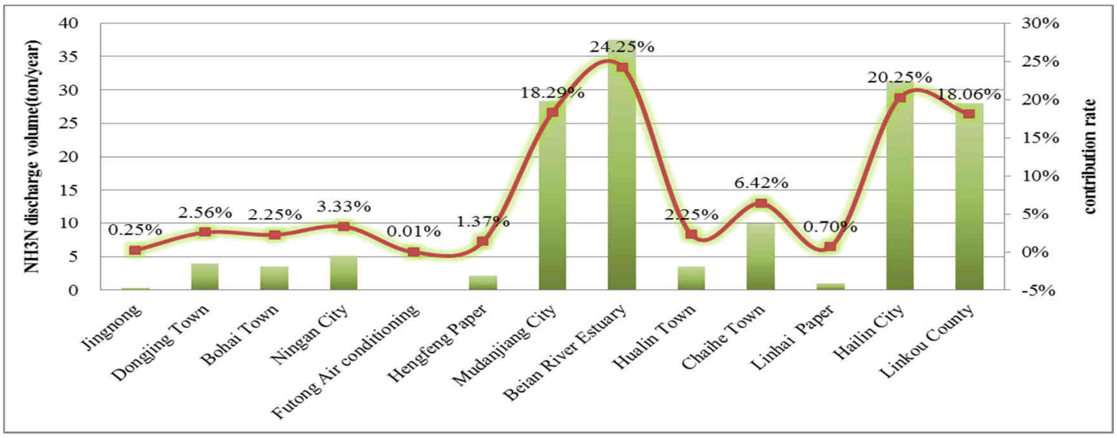


Figure 3-10. NH_3N discharge volume and contribution rate of each discharge outlet along Mudan River.

3.2.3 Current state of water quality analysis

(1) Interannual variability of water quality for trunk stream of Mudan River

On the basis of the percentage of section number for each water quality grade in trunk stream of Mudan River in the total number of evaluated sections, namely section water quality grade ratio method, has been adopted to make statistical calculation for various water quality grades in trunk stream of Mudan River from 2000 to 2014 (Table 3-1). Then inter-annual variability of percentage of each water quality grade has been analyzed (Figure 3-11).

Table 3-1 and Figure 3-11 shows that, from year 2000 to 2014, except for year 2008 in which grade III was dominated, the rest water grade of Mudan River were grade IV. From 2009 to 2013, grade III portion was higher, taking up over 1/3, indicating the water quality was improved. From 2013 to 2014, the water quality again dropped, especially in 2013, water of grade IV took up a percentage of 85.71%, reaching the peak within the fifteen years. Water of grade V in 2014 also took up 16.67% and water of inferior grade V (+V) occurred as well again after year 2010.

Table 3-1. Various water quality grades of trunk stream of Mudan River from 2000 to 2014 (%).

Grades	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
II	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
III	30	20	20	20	20	20	20	20	73.33	33.33	35.29	35.71	35.71	14.29	16.67
IV	60	70	40	40	50	80	50	60	26.67	55.56	41.18	42.86	57.14	85.71	61.11
V	0	10	10	30	30	0	20	20	0	5.56	17.65	21.43	7.14	0	16.67
+V	0	0	30	10	0	0	10	0	0	5.56	5.88	0	0	0	5.56

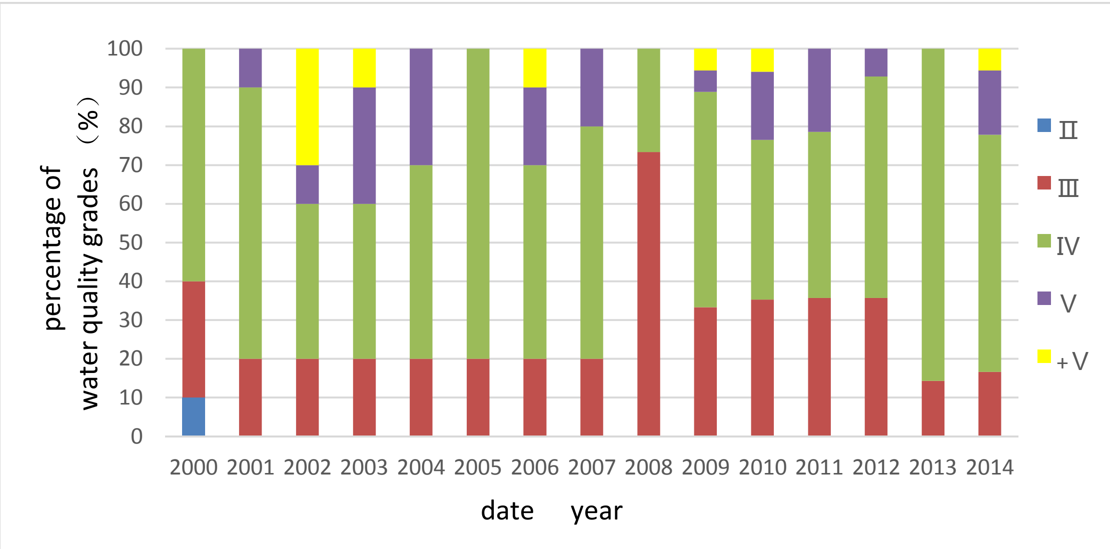


Figure 3-11. Interannual variability of percentages of various water quality grades in Mudan River.

(2) Spatial distribution of pollutants in trunk stream of Mudan River

Figure 3-12 presents that COD concentration of individual sections in different water periods along the river showed a small fluctuation in 2014. Except for Dashanjuzi and Mudan River estuary whose section concentrations demonstrated a large difference, concentrations of other sections in different water periods showed a small difference. Specifically, COD concentration peaked in Dashanjuzi section in high flow period, followed by Hualin Bridge and Chaihe Bridge; COD concentration peaked in Hualin Bridge section in middle flow period, followed by Chaihe Bridge, and Mudan River estuary was the last one. COD concentration in low flow period demonstrated an inverted "V" shape, peaking in Wenchun Bridge section and reaching the lowest point in Mudan River estuary section.

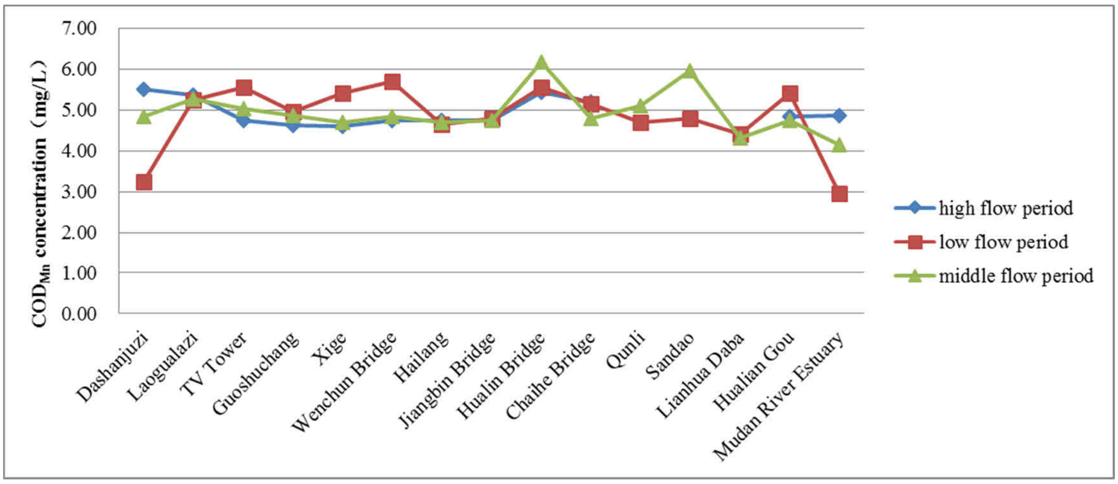


Figure 3-12. COD concentration variation of each section along trunk stream of Mudan River in 2014.

Figure 3-13 shows that NH_3N concentrations of individual sections in different water periods along trunk stream of Mudan River in 2014 showed a wavelike shape change, with small variations for different water periods. The changes of various sections in different water periods behaved in a similar way. Hualin Bridge section was the peak point for NH_3N concentrations in three water periods, followed by Sandao and Guoshuchang, while Dashanjuzi and Xige were the lowest.

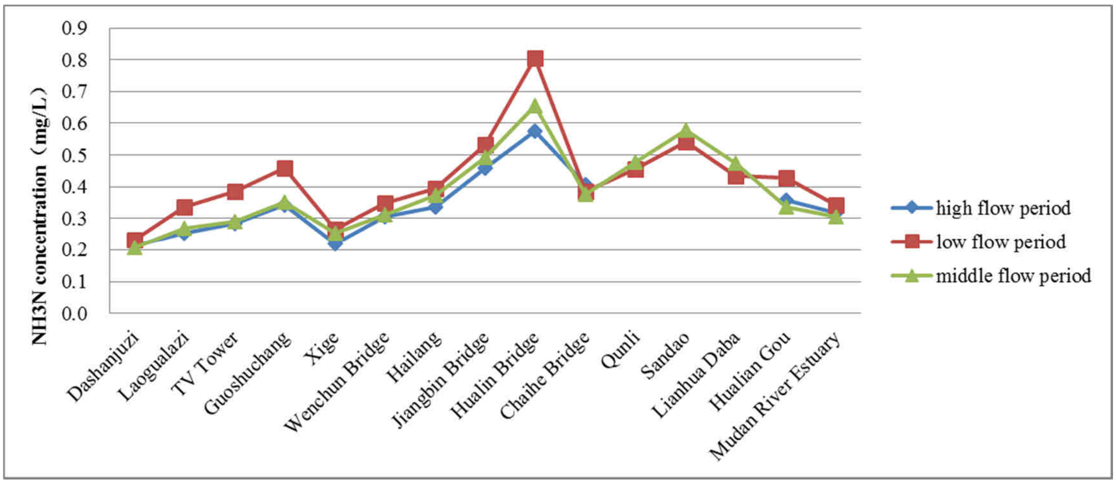


Figure 3-13. NH3N concentration variation of each section along trunk stream of Mudan River in 2014.

It can be found in Figure 3-14 that Total Phosphorous (TP) concentrations of individual sections in different water periods along trunk stream of Mudan River showed a wavelike change, with the concentrations in different periods being close. Concentration change trends in different water periods were similar. Hualin Bridge was the peak point for each water period, followed by Chaihe Bridge, Hualiangou and Xige. The rest sections were close in phosphorous concentration.

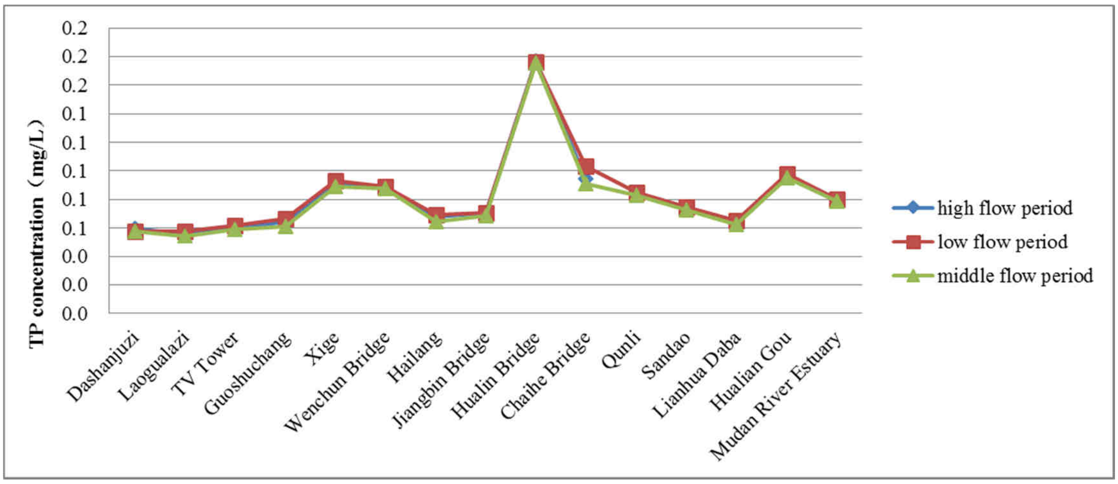


Figure 3-14. TP concentration variation of each section along trunk stream of Mudan River in 2014.

It's shown in Figure 3-15 that Total nitrogen (TN) concentrations of each section in different water periods along trunk stream of Mudan River showed a large variation. However, nitrogen concentrations in different water periods did not vary too much, with a similar change tendency. Hualin Bridge section was the peak point for TN concentrations of each section, while Jiangbing Bridge was the lowest.

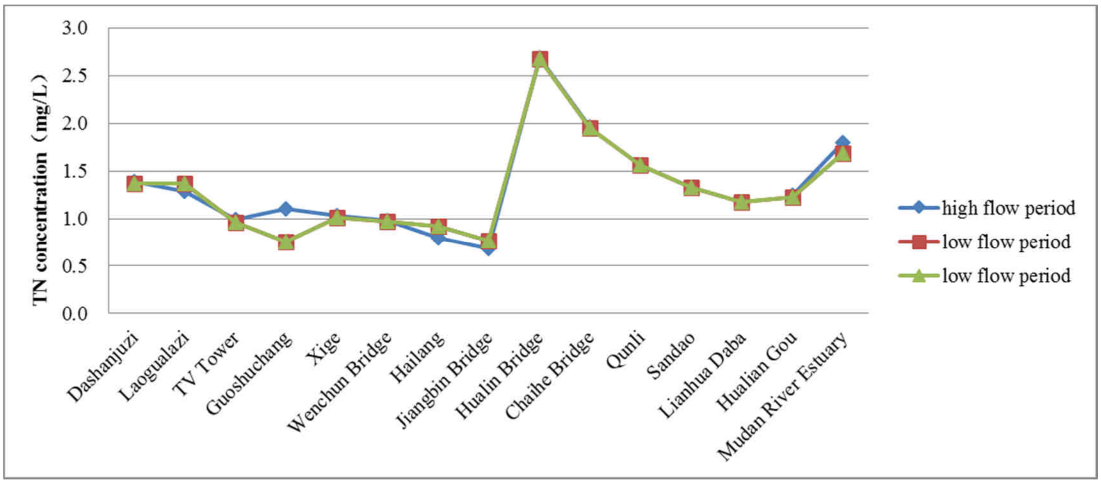


Figure 3-15. TN concentration variation of each section along trunk stream of Mudan River in 2014.

3.3 Modeling and verification for water quality of Mudan River

With preliminary understanding of the current water quality state and background information about Mudan River, the water quality model required for building a SEWS for Mudan River has been illustrated in this section, the flow diagram is shown in Figure 3-16. Firstly, determination of the river modeling segment and starting-ending time for simulation is needed. Then a grid model shall be constructed. Next, it needs to make the boundary condition scheme, and add the flows and pollutant concentrations of inlet, outlet, intake and tributary inlet to simulated river segment. Thirdly, it is required to make the pollutant load discharge scheme, input the discharge flow and pollutant load of waste water discharge outlet. Fourthly, the initial concentration of pollutant in each grid is required. Fifthly, after determination of the used model parameters according to the simulated scheme, it needs to create the initial data into required *.inp files, and start simulation program. Finally, read out of the simulation results from the output files of simulation to realize the visualization analysis and dynamic simulation of spatial-temporal change process.

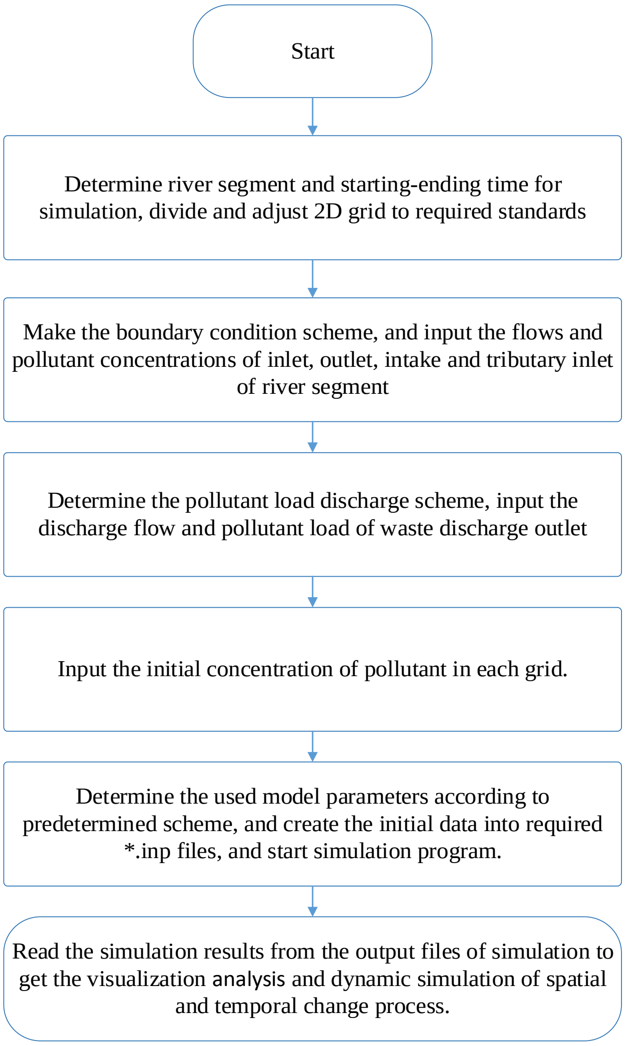


Figure 3-16. Model establishment for water quality of Mudan River and simulation flow.

3.3.1 Riverbed terrain

Underwater terrain for river course is one of the important factors influencing water flow and water quality, also one of the basic conditions for hydrodynamics and water quality model calculation. Measurements for underwater terrain of Mudan River between Xige section and Chaihe Bridge section have showed that terrain section of Mudan River almost presents parabolic shape, some shallow water exists on several wide sections. The change of riverbed elevation of measured river segment with the variation of longitudinal distance is described in Figure 3-17, which shows the elevation of riverbed generally drops from upper stream to lower stream, with an average slope of 0.0004. It can also be shown that adjacent sections in urban segment of Mudan River fluctuate a lot, showing an irregular variation, such complicated terrain has greatly challenged water quality numerical model simulation (Ma Yun et al., 2015).

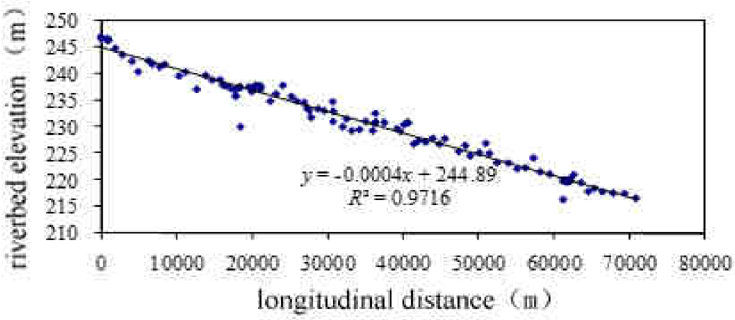


Figure 3-17. Variation of riverbed elevation along longitudinal distance.

3.3.2 Simulation range and time interval

Initial and terminal sections studied in this work are the Xige Section and the Chai River Bridge Section, respectively, covering a distance of 77.7 km. Hailang River, a main tributary of Mudan River, feeds into this river link, as shown in Figure 3-2. Considering that the river remains in a frozen state for five months every year, changes in sediment transport during its passage to ice-covered regime take place primarily due to changes in flow dynamics caused by the additional resistance at the upper boundary (Debolskaya et al., 2010), and its hydrodynamic parameters, comprehensive decay rate of pollutants and roughness in ice-covered period are significantly different from each other due to impacts of ice sheets. Therefore, simulations performed by this model are separately carried out for the ice-covered and open-water periods. For the purpose of data integrity of this study, the simulation interval for the open-water period is from 1 January 2012 to 15 November 2014; that is 1050 days in total. With regard to the model itself, January 1 of 2012 is set as the first day, and the number of days mentioned below is counted by starting from this date as the start. Ice-covered and open-water periods within the simulation interval are listed in Table 3-2.

Table 3-2. Simulation intervals for hydrodynamic water quality model

Simulation Interval (Days)	Number of Days	Water Season	Simulation Interval (Days)	Number of Days	Water Season
1–106	106	ice-covered period	107–335	229	open-water period
336–468	133	ice-covered period	469–698	230	open-water period
699–828	130	ice-covered period	829–1050	222	open-water period

3.3.3 Grid division and boundary condition setting

(1) Grid division

According to the actual terrain in the urban section of Mudan River trunk stream, orthogonal curvilinear grid is employed to divide its water body into 4,207 cells with a scale ranging between 24×43.6m and 176.9×241.3m (Figure 3-18). In addition, the grid matrix is 864 lines x 5 columns. Bottom elevation of the simulated river section lowers down from 245.7m on Xige section to 215.1m on Chai River Bridge section (Figure 3-19).

(2) Boundary conditions

Boundary conditions of the simulated flow include upstream on-coming flow of Xige Section, downstream water discharge from Chai River Bridge Section, flow from Hailang River fed into the trunk stream and quantity of wastewater effluent from sewage discharge outlets along the river. Within the simulated river section, as the only hydrology monitoring station named as S2 Station (Figure 3-2) is established on Hailang Section and water quality monitoring stations are set on Xige and Chai River Bridge sections, the on-coming flows for Xige and Hailang sections can be calculated on the basis of observed flow data in the S2 Station. According to the proportions taken by drainage areas controlled upstream of Xige Section and sub-stream of Hailang

River, respectively, observed flow data from the S2 Station are allocated to Xige Section and the mouth of Hailang River. Among them, flow of the former occupies 2/3 of that in S2 Station, while flow of the latter feeding into Mudan River accounts for 1/3 in that of such a station. Concerning boundary conditions of Chai River Bridge, the open boundary condition which is also known as the water level condition is adopted. Corresponding water level data can be acquired by using the observed water level in S2 Station minus the difference between the bottom elevation of the section and such an elevation of the Chai River Bridge section. It should be noted that pollution discharge at the sewage outlets of the trunk stream is observed flow data.

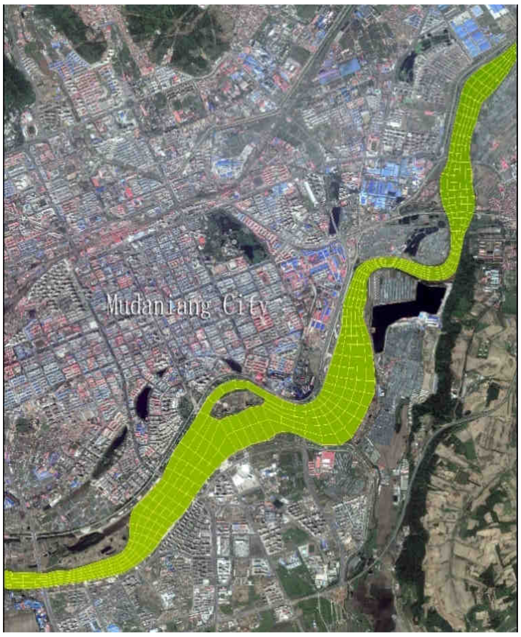


Figure 3-18. Grid division for urban section of the Mudan River trunk stream (part).

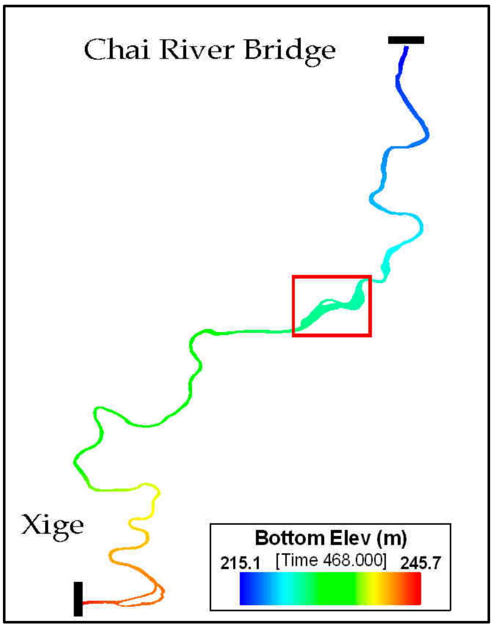


Figure 3-19. Bottom elevation diagram for urban section of the Mudan River trunk stream.

Boundary conditions for concentrations incorporate pollutant concentrations in upland water from upstream of Xige Section, in inflows from Hailang River and in all sewage outlets along the river. Four water quality monitoring sections such as Wenchun Bridge, Hailang, Jiangbin Bridge and Chai River Bridge are model verification sections. Monitoring water quality on section are carried out in January, February, May, June, July, August, September and October of each year; among them, January and February represent water qualities in ice-covered period, while the remaining months stand for those in open-water period. Within the trunk stream simulation section, 11 main sewage outlets are covered and their positions are shown in Figure 3-5.

According to the measured data, there is no continuous flow monitoring data from these 11 sewage discharge outlets, and only the total pollutant discharge of the whole year is available. Consequently, a constant value is adopted for the flow of those outlets. As the concentration of pollutants is monitored quarterly at each outlet, observed values are employed for boundary conditions of concentrations. Every year, sewage of 60 million m^3 approximately is discharged into the Mudan River trunk stream from those outlets; moreover, pollutants from some outlets are highly concentrated, even above the national sewage discharge standard and contaminated the river.

3.3.4 Model parameters and calibration method

$A_v A_b$ Model parameters that are required to be calibrated consist of horizontal diffusion coefficient, A_v (equation (1)-(2)), A_b (equation (7)-(8)), comprehensive decay rates of COD and NH_3N as well as bed roughness. Under conditions of ice-covered and open-water periods. Some researchers suggest that erosion rate is a critical index affecting the precision of water quality simulation (Li et al., 2014) ; but erosion rate is mainly influenced by particle size distribution, sediment bulk density, organic content, critical shear stress, and many external factors (i.e. flow velocity, wind speed, vegetation). According to research accomplishments given by Qu et al. (2013), there is little sediment runoff from the middle and lower reaches of the Mudan River trunk stream, and the river suspended load and the bed load are even less. For this reason, the proposed model did not take the influences of erosion rate on water quality into account.

Empirical value and trial methods can be brought together to define such parameters. In detail, empirical parameters calibrated with existing findings are firstly adopted and then adjusted, keeping the simulated results close to observed values; in the end, model parameters meeting simulation requirements are determined. Moreover, in order to guarantee model calculation convergence, the calculation step is set as 6s.

(1) Roughness

Roughness is one important parameter for resistance. As for urban segment of Mudan River, air temperature in low flow period is rather low and river may be ice-covered (Figure 3-20). Therefore, when analyzing water flow

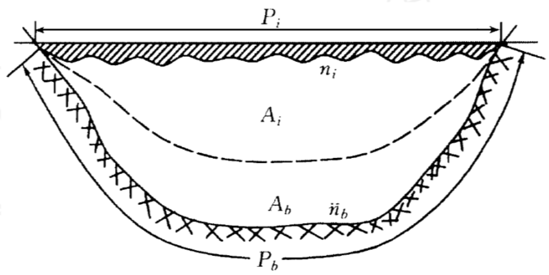


Figure 3-20. Cross section of frozen river course.

and water quality variation rule of this segment, riverbed roughness should be accounted, ice cover roughness also should be considered in low flow period. Water flow change plays a small role in water depth as for riverbed roughness, which could be taken as a constant in different water periods. Ice cover roughness only exists in low flow period, during which, water flow is relatively stable and the frozen mode is flat. Therefore, average roughness (0.01) of ice cover in early winter, middle winter and late winter has been taken for the ice cover roughness of simulated Mudan River segment in calculation. The comprehensive roughness in low flow period is calculated in accordance with equation (3-1).

$$\frac{n}{n_b} = \frac{0.63\left[1+\frac{P_i}{P_b}\right]^{5/3}}{1+\frac{n_b}{n_i}\left(\frac{P_i}{P_b}\right)^{5/3}} \qquad (3\text{-}1)$$

In which, n, n_i, n_b are comprehensive roughness, ice cover roughness and riverbed roughness, respectively. P_i, P_b are thicknesses of ice cover and riverbed, respectively.

Under-ice roughness indicates the resistance of ice cover to water flow. Different freezing conditions may cause huge difference in ice cover roughness. Slow river is often ice-covered due to thermal factors, leading to smooth under-ice surface and a low roughness. It was about 0.01-0.012 in early ice-covered period, and 0.008-0.01 in both stable ice-covered period and late ice-covered period. As for the ice cover roughness changes to a large extent between 0.01-0.1 depending on specific condition. Roughness adopted in this research is from the research of Wang Mei et al. (2013) by the above mentioned method, namely, roughness in ice-covered period is 0.43 and in open-water period is 0.035.

(2) Decay rate

Pollutant concentration will decay during transportation process by physical, chemical and biological interactions. The decay rate reflects how fast the pollutant is degraded in water. Currently, parameter needed to be calibrated in most water quality models is decay rate (Nares et al., 2013; Sperling et al., 2013; Wood et al., 2003). Guo (2008) concluded from researches on decay rate of COD and NH_3N in part of China rivers that: the decay rate of COD is 0.009~0.470/day, of NH_3N is 0.071~0.350/day.

Change of decay rate with water temperature is expressed in Arrhenius equation (3-2):

$$K_t = K_{20}\theta^{T-20} \quad (3\text{-}2)$$

In which, Θ is temperature coefficient, K_t degradation coefficient in question; K_{20} degradation coefficient in 20℃, T water temperature (Ma Yun et al., 2015).

Due to influence of meteorological factors, water temperature also demonstrates obvious seasonal changes. According to the actual conditions of urban section of Mudan River, water temperature is taken as 26.2℃ for high flow period, while 17.1℃ for middle flow period. In low flow period, because the water is ice-covered, theoretical water temperature should be 0~4℃. However, due to the shallow water of Mudan River, water temperature in such period is calculated as 0℃.

According to simulation intervals defined in Table 3-2, transport process simulations were performed for COD and NH_3N in simulated river reaches in ice-covered and open-water periods. Decay rates of COD and NH_3N were 0.03/day and 0.05/day in the open-water period, and 0.01/day and 0.02/day for the ice-covered period. Compared with the decay rates of other rivers in China, mean values for COD and NH_3N were 0.32/day and 0.23/day, respectively; COD and NH_3N decay rates in the Mudan River are thus at a lower level during both the ice-covered and open-water periods; especially the decay rate of NH_3N, which is even smaller than the minimum value (0.071/day) summarized by Guo et al. (2008). Considering possible reasons, they are related to the geographic position of the Mudan River which is further north than other rivers and also has a much lower multi-year average temperature (Weiler, 2011; Stratton, 1969).

(3) Other parameters

In addition to the roughness coefficient and decay rate, other parameters that need to be calibrated for the hydrodynamic process consist of the horizontal diffusion coefficients, Vertical turbulent viscosity and Vertical turbulent diffusivity. After the findings of Chen et al. (2007), Long et al. (2002), Huang et al. (2008) and HydroQual Inc. (2002) values of these three coefficients of 1.0 m^2/s, 1.0 × 10-7 m^2/s and 1.0 × 10-8 m^2/s, respectively, were used as references in this model.

3.3.5 Validation of hydrodynamics

As important parameters to characterize the size of resistances borne by river water bodies and one of the essential parameters for stream simulation and calculation, roughness coefficients reflects the impact of bed roughness on flow action. Under the condition of non-uniform flow in natural river courses, it is a coefficient under the combined actions of factors including stream equilibrium shape, watercourse hydraulic factor, section geometry and morphology, and bed surface characteristics and composition. The calculation of watercourse roughness mostly takes advantage of empirical or semi-empirical formulae. For watercourses in the open-water period, the roughness is computed by mainly taking the riverbed impacts into account. However, with respect to ice-covered rivers under the impact of ice sheets, boundary conditions for their free water surface turns into the ice sheet solid wall boundary condition; as vertical distributions of velocity are changed by ice sheets under the action of roughness, the velocity correspondingly decreases with the increase in watercourse roughness.

According to the simulation intervals defined in Table 3-2, hydrodynamic processes in ice-covered and open-water periods are simulated for the urban sections of the Mudan River trunk stream; furthermore, water level simulated results and observed water levels for the section at the S2 Station were compared with each other, as shown in Figure 3-21.

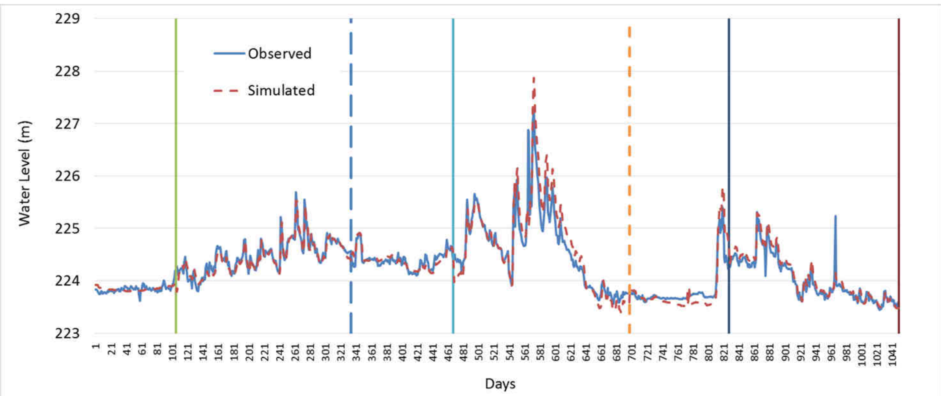


Figure 3-21. Result comparison between observed value and simulated value

The figure suggests that during both the ice-covered and open-water periods, the model calculation results are in excellent agreement with the observed values, which indicates the water level variation process of simulated river are significant. During the ice-covered period, the measured mean water level is 224.04 m and the corresponding depth of water is 0.90 m; then, the depth of water corresponding to the simulated mean water level of 224.03 m is 0.89 m; the average error is only 0.01 m and the relative error of the average water depth is 1.11%. In comparison, during the open-water period, while the measured mean water level was 224.39 m with a corresponding depth of water of 1.25 m, the water depth corresponding to the average simulated water level of 224.42 m is 1.28 m, their average error is 0.03 m and their relative error of the average water depth is 2.40%. On the one hand, according to simulated velocity results for different water seasons, the velocity in the open-water period is significantly greater than that in the ice-covered period because the inflow in the open-water period is larger than that in the ice-covered period.

Figure 3-22 reflects such a physical truth excellently. On the other hand, as observed from the aspect of flow field distribution characteristics, as the watercourse gently descends along the Xige section to the Chai River Bridge, the water flow direction is basically vertical to the watercourse section. Next to the center island of the watercourse, the flow direction changes due to its obstruction (Figure 3-22). Figure 3-22 shows the flow field simulated results on the 187th (high flow period) and 370th (dry season) days for the local channel in the research area. In this figure, the river widths of the #1 and #4 sections are 320 m and 181 m, respectively, and their flow velocities are correspondingly 0.43 m/s and 0.51 m/s on the 187th day while they are 0.29 m/s and 0.35m/s on the 370th day. Concerning the #2 and #3 sections, their river widths are 813 m and 903 m; and the flow velocities on the 187th and the 370th days are 0.19 m/s and 0.12 m/s, and 0.14 m/s and 0.09 m/s. These results indicate that the flow velocity of the narrow reach is faster than that of the wide one, this is in conformity with physical truth of the research area.

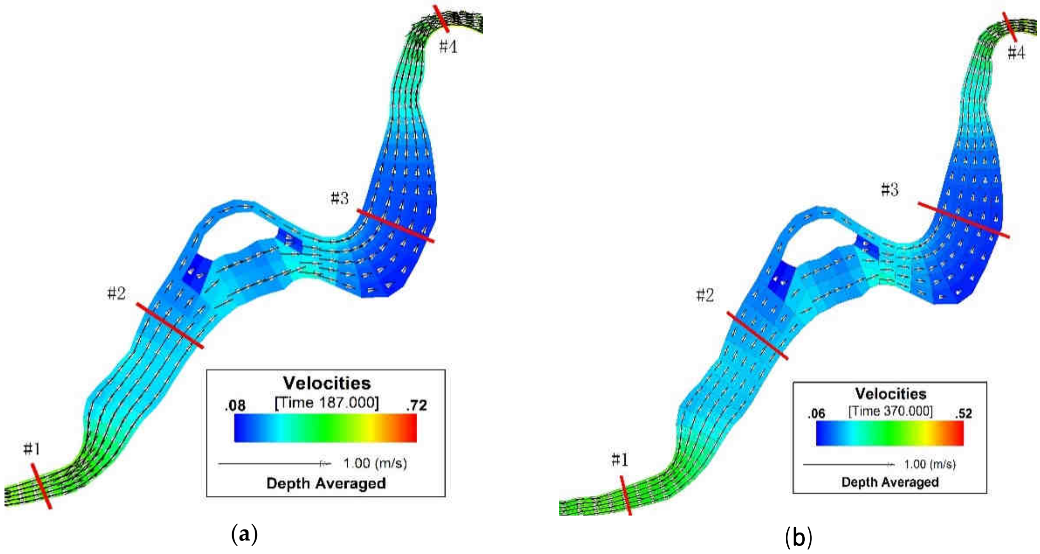


Figure 3-22. Schematic diagram for flow field in the urban section of Mudan River trunk stream. (a) open-water period; (b) ice-covered period.

3.3.6 Validation of water quality

(1) Parameter calibration result analysis

In the transport process during which pollutants enter into the river, concentration decay takes place by physical, chemical and biological actions, and the decay rates reflect the degradation velocities of those pollutants under the action of the water body. Currently, the parameter that needs to be calibrated by the overwhelming majority of water quality models is decay rate (Nares et al., 2013; Wool et al., 2003; Sperling et al., 2013). Through summarizing the findings of COD and NH3N decay rates in some rivers in China, it has been concluded by Guo et al. (2008) that the COD decay rate for Chinese rivers is 0.009–0.470/day and it is 0.071–0.350/day for NH3N. Most of such rivers are however situated in warm areas with no or short ice-covered period.

According to the simulated results, decay rates of COD and NH3N in ice-covered period are both smaller than those in the open-water period. The main reasons are as follows: first, the water temperature in the ice-covered period is relatively low; and extremely low temperatures can directly affect the degradation of pollutants by microorganisms (Druon et al., 2010). Second, due to the decreased inflow in the ice-covered period, the dilution effect of rivers on pollutants drained into them weakens; furthermore, together with the influence of ice sheets, the fluidity of the water body becomes poor so as to impact physical, chemical and biological reaction processes of pollutants (Wright, 1979; Pu et al., 1999). Third, the water body is isolated from the atmosphere by ice sheets during the ice-covered period; consequently, reoxygenation due to natural aeration almost completely stops and the concentration of dissolved oxygen remains in a low state so that sources of dissolved oxygen required by organic matter degradation are restricted followed by a drop of the degradation rate (Wang et al., 2003).

(2) Simulated Result Analysis

Figure 3-23 compares the observed value and the simulated value of COD in the Mudan River trunk stream. It shows that variation trends of those two values are in good agreement. Statistical analysis results for the simulated and observed values of COD are shown in Table 3-3.

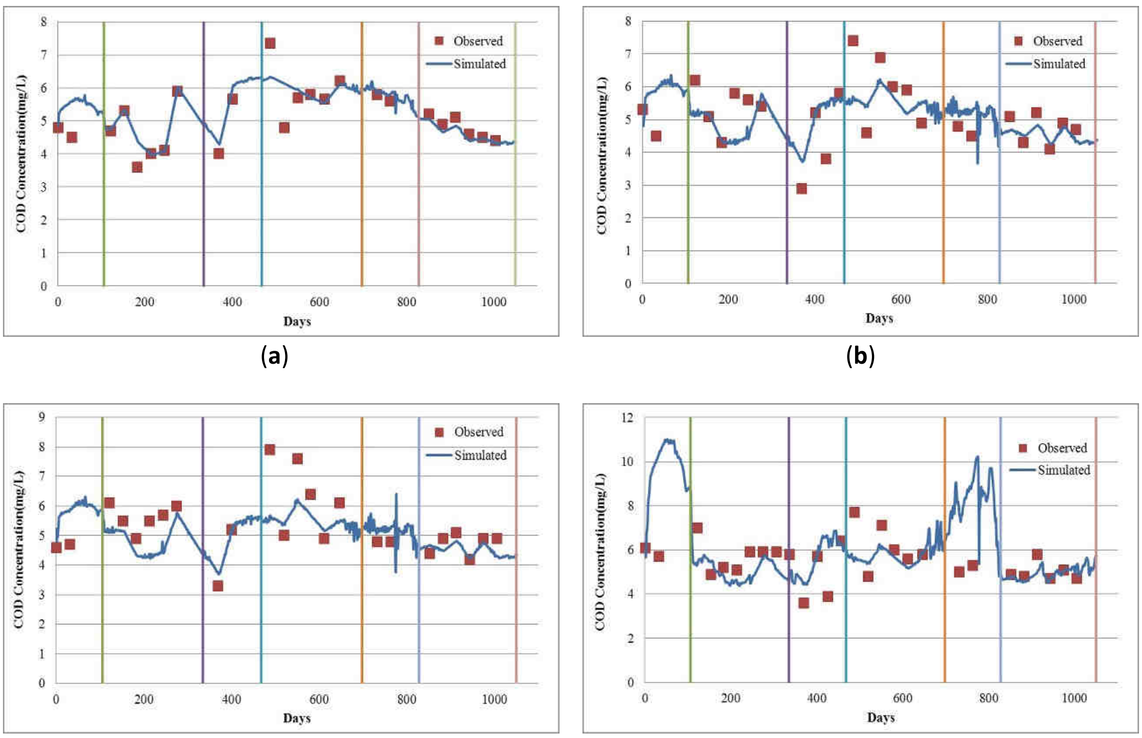


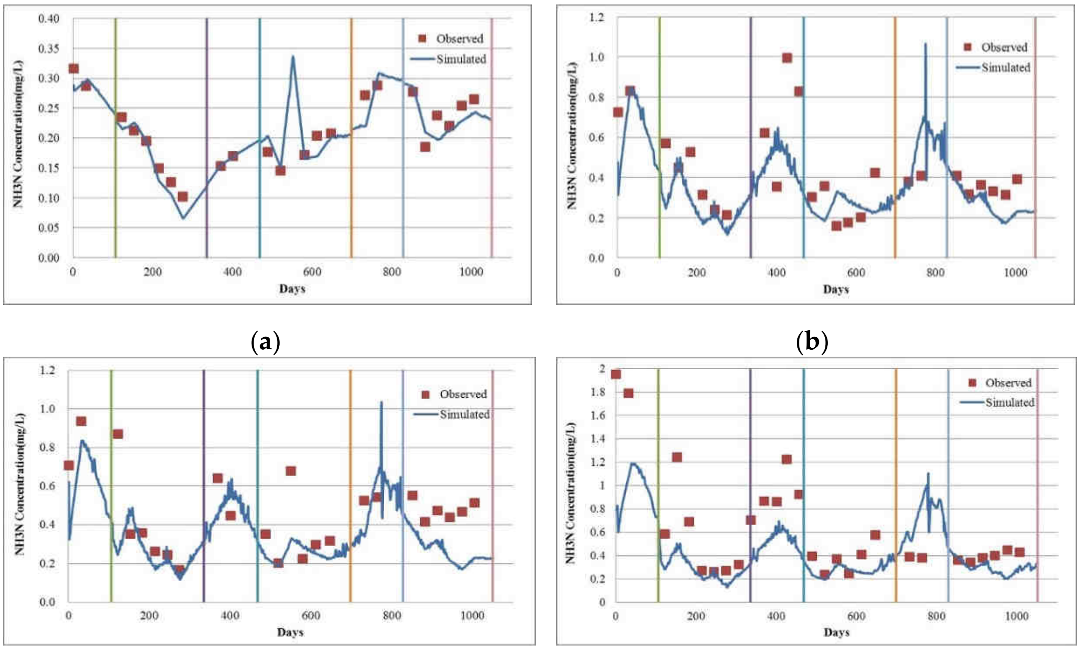
Figure 3-23. Result contrast charts for observed and simulated values of COD in the Mudan River trunk stream. (a) Wenchun Bridge; (b) Hailang; (c) Jiangbin Bridge; (d) Chai River Bridge.

Table 3-3. Statistical results for COD simulated and observed values in the Mudan River trunk

Simulation Period	Wenchun Bridge		Hailang		Jiangbin Bridge		Chai River Bridge	
	Sample Size	Average Relative Error (%)	Sample Size	Average Relative Error (%)	Sample Size	Average Relative Error (%)	Sample Size	Average Relative Error (%)
ice-covered period	6	7.19	8	17.44	6	11.53	9	35.41
open-water period	18	5.42	18	10.66	18	11.07	19	10.39
in total	24	5.86	26	12.75	24	11.18	28	18.43

On the whole, from the perspective of the four verified sections, the Chai River Bridge one has the maximum average relative error of 18.43%, while the Wenchun Bridge section has the minimum (5.86%). Based on different simulation periods, the simulation effects of the four sections during the open-water period are better than those in the ice-covered period. To be specific, the simulation errors for the Wenchun and Jiangbin bridge section in both the ice-covered and open-water periods were 7.19% vs. 5.42% and 11.53% vs. 11.07%, respectively; in comparison, those for Hailang and Chai River Bridge in these periods were greatly different from each other, especially the average relative error of the latter in the ice-covered period is significantly larger than that in the open-water period (35.41% vs. 10.39%). The main reason causing such differences was insufficient observed values for the ice-covered period (Table 3-2). Moreover, the lack of concentration information in unmeasured months can exert an influence on the simulation effects of the ice-covered period.

As shown in Figure 3-24, a comparison of observed and simulated NH_3N values in the Mudan River trunk stream, the NH_3N simulation result can also be utilized to roughly reflect the actual variation situation. Table 3-4 presents the statistical analysis results for the simulated and observed values of NH_3N in the Mudan River trunk stream model. As a whole, for the four verified sections, the Chai River Bridge has the maximum average relative error of 39.58% while that of the Wenchun Bridge is the minimum (14.88).



(a) (b) (c) (d)

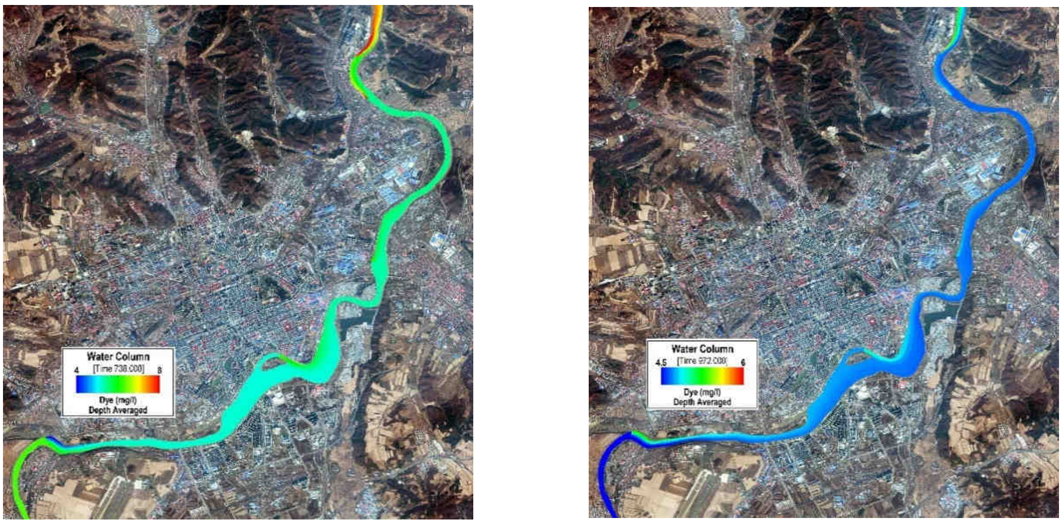
Figure 3-24. Result comparison for observed and simulated values of NH_3N in the Mudan River trunk stream. (a) Wenchun Bridge; (b) Hailang; (c) Jiangbin Bridge; (d) Chai River Bridge.

Table 3-4. Statistical analysis for simulation and observed values of NH_3N in the Mudan River trunk stream model.

Simulation Period	Wenchun Bridge		Hailang		Jiangbin Bridge		Chai River Bridge	
	Sample Size	Average Relative Error (%)	Sample Size	Average Relative Error (%)	Sample Size	Average Relative Error (%)	Sample Size	Average Relative Error (%)
ice-covered period	6	10.92	8	35.94	6	21.66	9	55.15
open-water period	17	16.28	17	33.42	18	35.35	19	32.32
in total	23	14.88	25	34.23	24	31.93	28	39.58

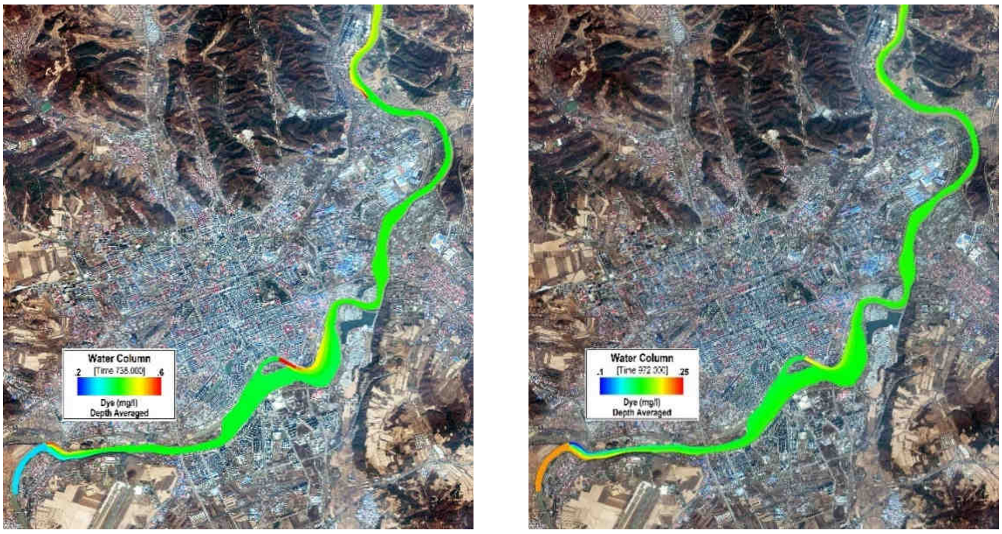
On the basis of the different simulation periods, unlike the COD simulation effect, there are differences in the simulation effects of these four sections during the open-water period. Among them, the simulation errors for the Wenchun and Jiangbin bridge sections in the ice-covered period are smaller than those in the open-water periods. However, such effects for Hailang and the Chai River Bridge in the open-water period are better than those in the ice-covered period. In general, it is feasible to apply the NH3N model to the Mudan River trunk stream despite the fact that its simulation accuracy is not as high as that of the COD simulation effects. The main cause for this is that the COD pollution source is primarily industrial discharge; continuous monitoring data are recorded for the concentration and discharge capacity of such pollutants and those data are able to support the model simulation. However, its NH_3N pollution sources are mainly non-point source pollution and sanitary sewage within the drainage basin (Wang et al., 2011). As only monitoring data of the sanitary sewage are noted down for the Mudan River trunk stream, which is lacking in non-point source pollution monitoring data, the accuracy of boundary conditions is lowered when it comes to the concentration of NH_3N, so as to further exert an impact on the precision of our model simulation.

Figure 3-25 and Figure 3-26 refer to the spatial distribution diagrams of the COD and NH3N simulated results in the ice-covered and open-water periods. On their basis, it can be clearly observed that as the concentration of pollutants discharged into the Mudan River from sewage outlets is intensively mixed after a certain distance under the action of attenuation and dispersion, sectional concentrations basically become consistent (Fischer et al., 1979). Basically a mixed concentration of COD or NH_3N can be attained about 3 km to 5 km away from the sewage outlets where they are discharged. In case that a segment of watercourse in the downstream of sewage outlet is relatively straight, basically the mixed concentration can be reached after a long distance; while when such a segment undergoes twists and turns, the concentration can be mixed within a short distance under the impact of turbulence.



(**a**) (**b**)

Figure 3-25. COD concentration distribution in the Mudan River trunk stream (urban sections).



(**a**) (**b**)

Figure 3-26. NH_3N concentration distribution in the Mudan River trunk stream (urban sections). (a) 1-7-2014 (ice-covered period); (b) 8-29-2014 (open-water period).

3.4 Problem analysis

Although simulated values of COD and NH_3N display excellent consistency with their observed values, certain errors still exist. It has been found that the main reasons leading to pollutant concentration calculation errors included the following aspects:

(1) Error of bottom elevation generalization

With complex sectional form, a wide Mudan River reach often covers flood plains with different depths; therefore, its sections are usually in the shape of letters U or W. Given the stability of the model calculation, a watercourse was generalized as a rectangle in this model. In other words, the riverbed was deemed to be horizontal within cells. Due to such a generalization, some river sections, especially the W-shaped ones, can be distorted within a cell leading to simulation errors.

(2) Variability of pollution sources discharged into the watercourse from urban storm drainage pipe networks and sewage pipes

Within the range of the simulated river reaches, there are large cities such as Mudanjiang City, Ning’an City and Hailin City, and some villages and towns including Wenchun Town, Hualin Town and Chaihe Town and so on are situated there too. Furthermore, along the river, there are many rainwater outflows and pollution discharge outlets whose discharge capacities and sewage concentrations mostly change dynamically. Especially the sanitary sewage outflow, and water discharges varied enormously during the different seasons. In this study, the adopted observed pollutant concentrations and wastewater discharges were measured quarterly. However, as data of some outflows were not monitored, the accuracy of the model flow and concentration boundary conditions declined.

(3) Dispersed non-point agricultural source pollution on each side of the trunk stream and relevant measurement difficulties

Water quality pollution in the Mudan River watershed is caused by the joint action of the point source pollution and the non-point source pollution (Chui et al., 2005). Non-point source pollution sources within the Mudan River watershed mainly include urban non-point source pollution, irrigation backwater pollution, livestock breeding pollution and garbage pollution piled up disorderly in a rural area. In particular, the urban non-point sources refer to contamination of surface water, underground water and soil, caused by sanitary sewage and a small amount of industrial wastewater which are discharged into the river via a natural stream, an artificial stream or dispersed drainage; and such sanitary sewage and industrial wastewater may occur in multiple medium-and-small-sized towns and vast rural areas where no sewage treatment plants are established in the Mudan River region. Along the Mudan River, there are masses of farmlands. During irrigation, flood irrigation is adopted in most case so that a good deal of pesticides, chemical fertilizers and soil organic matter enter the river along with the irrigation backwater. In addition, there is large-scale livestock and poultry breeding along the Mudan River, but pollution prevention measures are inadequate among most of the breeding industries. Consequently, faeces produced by livestock breeding are piled up at random and it is extremely easy for them to be washed away in huge quantities. Especially during the rainy season, part of them seep into the underground and may contaminate the ground water; and the other part can flow into nearby rivers, lakes and reservoirs along with rainwater resulting in organic pollution of the surface water. Moreover, sufficient waste treatment plants have not been constructed in different towns along the Mudan River watershed. Garbage is randomly piled up on both sides of the watercourse; after they enter a water body due to rainwash, and it can be contaminated directly. Considering that those pollution sources are very scattered and hard to control, parameter calibration and verification of the model is very challenging.

3.5 Chapter conclusion

The construction of water quality model of Mudan River is the basis of SEWS, which needs to think about the specific situation of research objective. This chapter introduced the watershed and analyzed the water quality status of Mudan River. According to the topography data and water quality monitoring data of the watercourse, this chapter detailed how to make the river grid system and set up the model boundary condition, a new 2D hydrodynamic water quality model has been established to simulate COD and NH_3N concentrations during the ice-covered and open-water periods for the Mudan River in China. Moreover, we calibrated and verified the involved parameters, including the dispersion coefficient, the roughness and the comprehensive decay rate. On the basis of our research and analysis, some conclusions can be drawn as follows: 1) Research findings show that the concentration simulation errors of COD in the four sections for model verification range from 5.86% to 18.43%; while for those of NH_3N, are between 14.88% and 39.58%. 2) The decay rate of COD and NH_3N during the ice-covered period is lower than that in the open-water period. According to the research results, in the trunk stream of the Mudan River, the decay rates of COD and NH_3N during the open-water period are 0.03/day and 0.05/day while they are 0.01/day and 0.02/day during the ice-covered period. 3) In this research, the roughness adopted for the ice-covered period was 0.043 and 0.035 for the open-water period. Obtaining favorable simulation effects indicates that these two parameters were selected appropriately. 4) Comparing with the decay rates of other rivers in China, those of COD and NH_3N in the Mudan River are relatively lower. This may be due to the low annual average temperature of this river which is located in the cold north region.

As a whole, after the analysis and verification of the model, it's satisfied and can be coupled to GIS to construct the water quality SEWS.

Chapter 4

Coupling GIS distributed computation frame and file system of the water quality model

4.1 Integration of the water quality model with GIS

GIS is a powerful tool to manage and analyze the geographic information, which has such features as regionality, hierarchy, distributivity, dynamicality, and complexity. System development and integration has been the mainstream direction of GIS technology in recent decades. Integration of environmental models with GIS strongly depends on the functionality and complexity of the model. The following factors also shape some effects on integration with GIS: the requirement of the model on basic data and GIS functions, the practicability of the model's interface, the compatibility of data format, the hardware environment, and the system structure of the model software. Combination of GIS and EFDC has been used in simulating and forecasting water environment. Peng et al. (2010) established a GIS-based WASP model for water quality management to the Lower Charles River in order to improve data integration and management capabilities. By coupling the WASP with EFDC software, they found that this GIS-based WASP model can simplify data flow of water quality modelling and enhance the capability of data analysis for the management. Wang et al. (2006) established an integrative model to couple EFDC, WASP, and SWAT with GIS for Miyun Reservoir. The application range of EFDC software is not second to that of WASP model and is used in simulation and prediction of nutritive salt, algae and toxic substances (Ai Hainan et al., 2014). Both Peng et al. and Wang et al adopt a loose mode in their studies to couple the simulation model with GIS. According to the magnitude and the mode of integration, the approaches for coupling GIS with environmental models can be classified into the following 4 types (Zheng et al., 2005; Zhang et al., 2001; Nyerges et al., 1992; Wang, 2008).

(1) Standalone application

In different hardware environments, data exchange between different data models in GIS and the models usually is implemented manually (for example, ASCII file). Operation personnel play a vital role in GIS and the model interface while the requirement on the programming ability of operation technical personnel is not high.

(2) Loose coupling

This is a loose integration, namely, each integrated part is coupled by external interfaces (input and output files). Each part is not integrated together. The goal of integration is to utilize respective functions and the integrated approach realizes their data exchange by file exchange mechanism.

(3) Tight coupling

In this integrated approach, the data format of the water quality model still differs from that of GIS software. However, it can automatically execute bidirectional data access under the condition of no manual intervention. This integrated approach needs more programming tasks and operation personnel are still responsible for data integration.

(4) Complete integration

In the complete integration system of GIS and the water quality model, the water quality model and GIS model belong to different modules of the same integrated system. Data access is based on the same data model and share the same data management system. The interaction between subsystems is very simple and effective. However, the software development efforts and difficulty of this integrated approach are very huge.

Due to speciality and complexity, the water quality model based on mechanism is an integral module. If it is completely reconstructed under GIS support, the workload is great. Thus, it is not suggested disassembling because of the different requirements of the information system. When constructing Mudan River SEWS, this system considers the following criteria in the design of the water quality model: it has good scalability and integration, maintainability, extendibility, function completeness and usability. Therefore, according to the characters of the water quality model and the elastic demands of the system architecture, tight coupling is adopted. This approach can reduce the coding workload of integration development and meanwhile guarantee the independent integrity of the water quality model and make it easy to update and maintain.

EFDC software code adopts structured program language Fortran. On the whole, its file flow can be classified into three parts: input file, output file and master control file. Input files include topographic data files, model parameter data files and time series data files. Master control files control whether activating some functions of the model or not, set the source data (for example, defining the number of inflow river channels and the number of sewage outlets), control output variables and frequency of the model results, set the position of open boundary, and setting the simulation periods. Output files mainly are used to store the computing results and relevant intermediate achievements of output variables under spatial-temporal dimension (Jia Peng et al., 2015).

Codes of Mudan River water quality model can be programmed by Intel Parallel Studio XE 2011 combined with Microsoft Visual Studio 2010. After compiling source codes, EFDC.exe is the executable program of the software, executing the simulation computing tasks. Input files are constituted of master control files EFDC.inp and a series of other .inp auxiliary files, mainly used for the model function selection and setting initial conditions and boundary conditions. Output file .out is used to store the simulation results of the model, for example, flow velocity, flow rate and pollutant concentration. Integration of GIS and water environment dynamic model means the comprehensive integration and application of GIS functional components, the executable programs, input files and output files.

Table 4- 1. Main input and output files of two-dimension Mudan River water quality model (Li, 2009).

File name	File type	File function
EFDC.exe	executable program file	Assign the screen output control file of cells in the operation process of the model.
EFDC.inp	master control file of the model	This file includes the description function of operation control parameters, output control parameters, and model physical information. It is a master control file of Mudan River water quality model.
cell.inp	cell information file	Digitize the outline of water body. All cells are given integer variables to represent types. For example, 5 represents lake surface, 0 represents land, and 9 represents land-water area. Program computing recognizes water body or land according to different values.
dxdy.inp	cell information file	Assign horizontal cell spacing, water depth, reservoir bottom elevation, reservoir bottom roughness and vegetation types.
lxly.inp	cell information file	Store grid center point coordinates and rotation matrix
corners.inp	cell information file	Store cell center point coordinates and four-angle coordinates
qser.inp	time series file	Store time series of flow rate
dser.inp	time series file	Store time series of pollutant concentration
dye.inp	initial concentration file	Store pollutant initial concentration
show.inp	Operation display	display control files in the program execution

	control file	process
DYECONH.OUT	output files of simulation results of water quality	simulation results of pollutants
UUUDMPF.ASC	output files of simulation results of flow velocity	simulation results of flow velocity in direction X
VVVDMPF.ASC	output files of simulation results of flow velocity	simulation results of flow velocity in direction Y
DYEDMPF.ASC	output result files of pollutant concentration	simulation results of pollutants (which are consistent with the output results of DYECONH.OUT)
SELDMPF.ASC	output result files of water level	simulation results of water level

The interactive contents of GIS and Mudan River water quality model finally reflects in the schemes of daily and emergency water quality simulation. Difference between different water quality simulation schemes based on the water quality model also is reflected in this file. Thus, the core content of distributed file process is the water quality simulation scheme files.

4.2 Analysis of the characteristics of model computing and distributed architecture design

EFDC software adopts a standalone-version single-thread computing simulation approach. For large data-volume hydrodynamic and water quality simulation process, it is an inefficient operation mode. For Mudan River water quality model constructed in Chapter 3, the simulated two-dimension water quality area is a part of Mudan River trunk stream. This model owns 4,207 grids and the step length is 6s. Relatively, the analog quantity of this model is very small, but completing one single-scheme simulation costs about 108.2 minutes in the environment of desktop workstation of windows 7 64bit, Intel (R) Core (TM) i7-2760QM CPU@2.4GHz 2.4GHz, 10G RAM. EFDC usually is used in the case with greater construction scope and more grids, which needs more time. If multiple model schemes are run at the same time and the parameters of these models are supposed the same, as well as the simulation regions, the total elapsed time is multiplying single-model scheme operation time by the number of simulation schemes, neglecting the initiating elapsed time of the model computing, shown as Figure 4-1. If the computing technology of the distributed model is adopted, the time-consuming computing can be changed, even if each single server only permit one simulation scheme and the uploading time of network transmission and scheme files and the downloading time of the simulation results are neglected (because these elapsed time is insignificant compared with the size of the distributed network and the elapsed time of the model computing. Especially, distributed server nodes usually are deployed in the same local area network. If the distributed file system is adopted, it is more obvious). Then, the elapsed time of operating 6 same simulation schemes on 6 same distributed server nodes is equal to the elapsed time of single server in the computing mode of standalone server. Meanwhile, when the number of simulation schemes increases, node servers can be dynamically expanded or decreased so as to increase or save server resources. Therefore, it is necessary to construct a distributed dynamic elastic computing system for the Mudan water quality model.

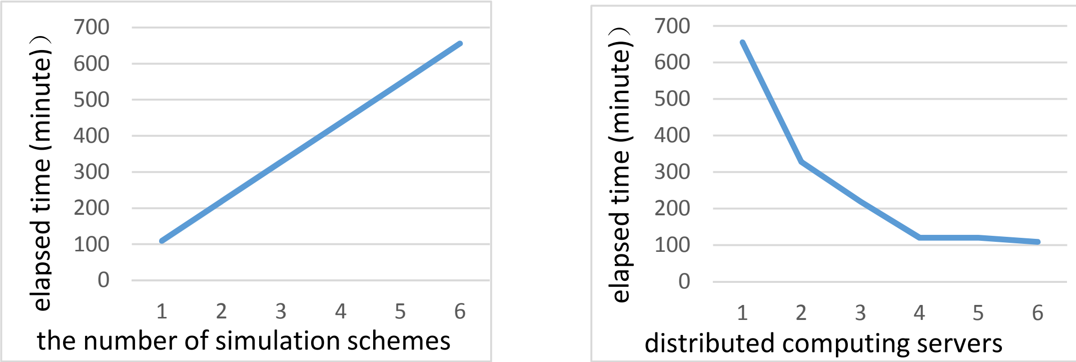


Figure 4-1. Elapsed time of 6 simulation schemes on single server and elapsed time of 6 simulation schemes on multi-server.

This study puts forwards a distributed computing method for the water quality model and its distributed architecture design is shown as Figure 4-2. The server is constituted of master server, model computing elastic nodes, database and distributed file system. Master server is in charge of completing model computing task scheduling, namely, by multiple-task scheduling mechanism, selecting relatively spare elastic nodes with the strongest processing capacity to feedback to the operation terminal of the client. The client deploys the SEWS of Mudan River used by business operators or research analysts. This system realizes functions of water quality data collection and storage, water quality data evaluation, water quality trend analysis, and water quality warning and simulation.

Elastic node cluster of the water quality model computing is constituted of the processing nodes of one or more model simulation schemes. Processing nodes of the model simulation scheme are responsible for completing the task request submitted by the client. Elastic nodes send current processing capacity information regularly to the master server so as to realize its load balancing scheduling. Elastic nodes are classified into entity elastic node server and virtual elastic node server, respectively referring to nodes in the physical environment and the node server in the virtual environment.

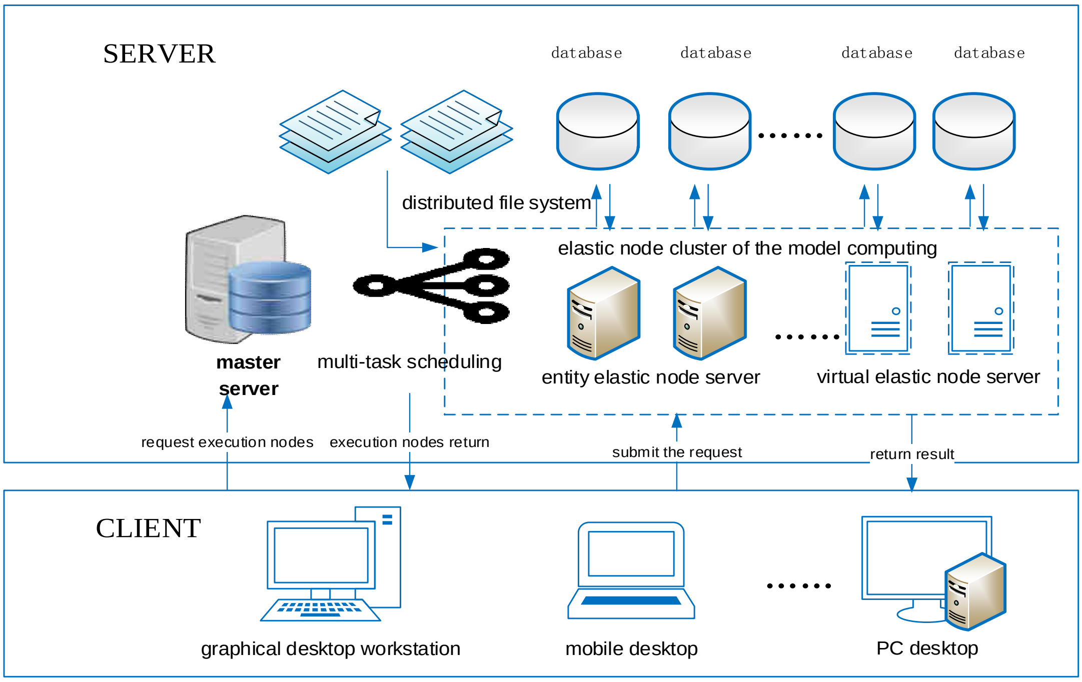


Figure 4-2. Overall architecture of distributed model computing.

4.3 Server design of distributed water quality model computing

The reasonableness of server design has a direct influence on its performance. The server design includes data package-based Socket Communication Server (SCS) and I/O Completion Port (IOCP).

(1)Data package-based SCS

To improve the communication efficiency between servers and client, socket communication mode is adopted for communication. Complete communication link is constituted of Client Socket and Server Socket, which adopt the same binary protocol communication. In binary communication, custom data package is used to package information content. Custom package is made up of head information and body content, shown as Figure 4-3.

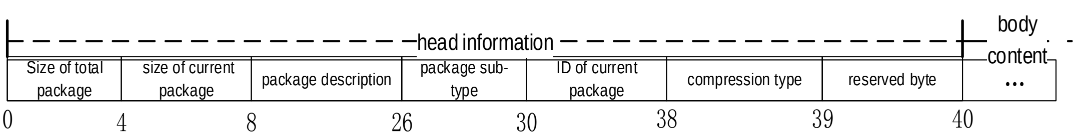


Figure 4-3. Structure of data package.

The total length of socket data acceptance is limited for one time and it is necessary to split and combine large packages. To identify each package, when the client encapsulates data package, unique ID with length of 8 bytes generates automatically and the subpackage owns the same ID as the total package. After the server accepts packet processing, feedback information also is encapsulated by data package. To make the client accurately identify the feedback information of the server, the server sets the feedback package with the same ID as reception package. In the interaction process with the server, 18-byte description character string and 4-byte subtype are utilized to classify the request contents of packages. Packages of different types represent different requests of the client. Data packages support data compression and 1-byte is used to represent the compression type of current package.

The body adopts flow mode to record data. Arbitrary number Byte, Short, Int, Long, Float, Double and String can be added in the package. For example, one floating-point number is added in the package, the length of the package increases by 4 bytes. When the package is interpreted, data in the package is read according to the same sequence (Chen Jun, 2012).

(2)IOCP-based server

Socket technology is utilized to realize the communication between the server and client. The server hosts the server socket and deals with the requests of the client. CPU executes the model computing processing and IOCP model design is adopted. IOCP model is one of the most complex I/O model. When one application program needs to manage numerous sockets at the same time, this model usually can realize the optimal system performance (Wu et al., 2007). Essentially, IOCP model creates IOCP object by scheduling Windows API function CreateIocompletionPort. By the thread with specified number, the thread requests on the blocked port are managed to provide services for the completed overlapping I/O request, shown as Figure 4-4.

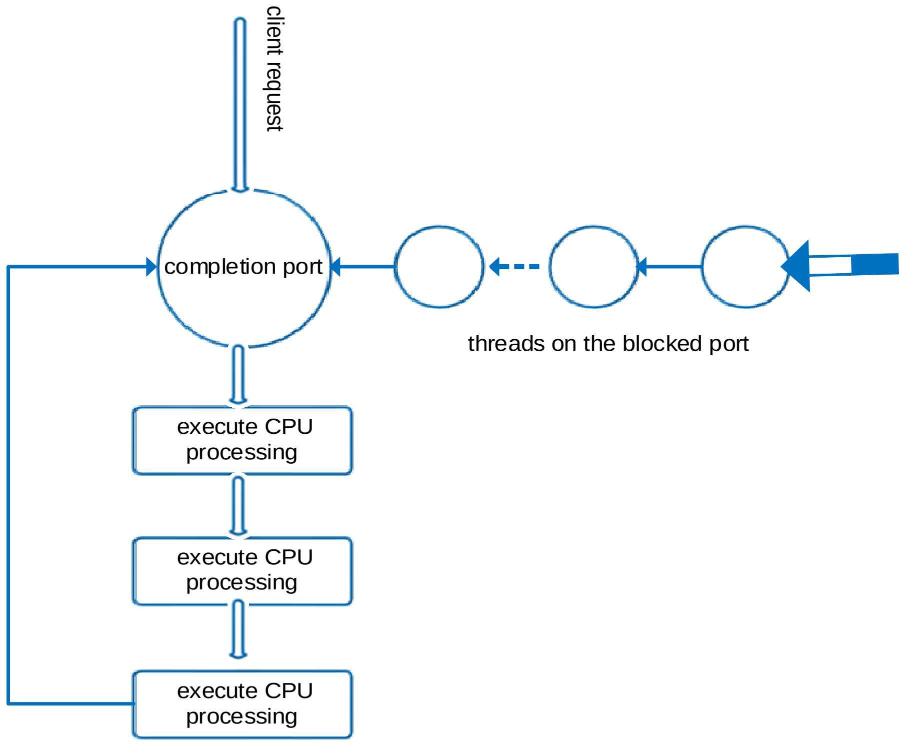


Figure 4-4. Completion port model (IOCP)-based server model operation.

Using this model needs, 1) calling Create CompletionPort function; 2) creating one I/O completion port; 3) creating a thread pool of mission computing; 4) cyclically scheduling function GetQueuedCompletionStatus to obtain I/O operation results by worker thread; 5) cyclically scheduling Accept function to wait for the link request of the client by the master thread; 6) returning Accept function of the master thread and creating a new link when the client requests link and associating socket handle with the completion port by AssociateDeviceWithCompletionPort function and using asynchronous WSASend and WSARecv to realize data delivery. The calls of WSASend and WSARecv will return immediately and actual processing is completed by Windows system; 7) the master thread continues the next cycle and prevents the new client link waiting at Accept function; 8) asynchronous WSASend and WSARecv is utilized to receive data and send the results to the completion port; 9) worker thread GetQueuedCompletionStatus function returns immediately and the results of WSASend and WSARecv operation are obtained from the completion port; 10) the project thread processes the result data and then

WSASend and WSARecv are called to continue the next cycle. Shown as Figure 4-5, dotted line represents the processing of Windows system, which needs program intervention.

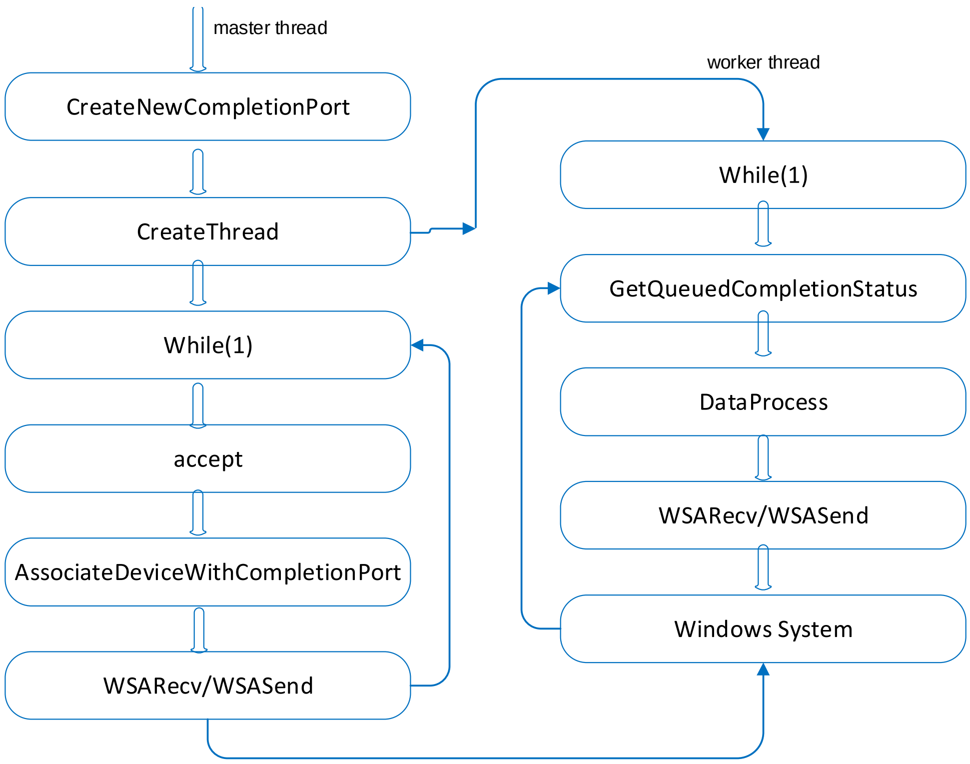


Figure 4-5. Processing flow of I/O completion port.

4.3.1 Communication architecture design of the simulation server cluster

In the distributed server cluster of the water quality model, there are two kinds of servers, master server and elastic execution nodes. The communication architecture is shown as Figure 4-6.

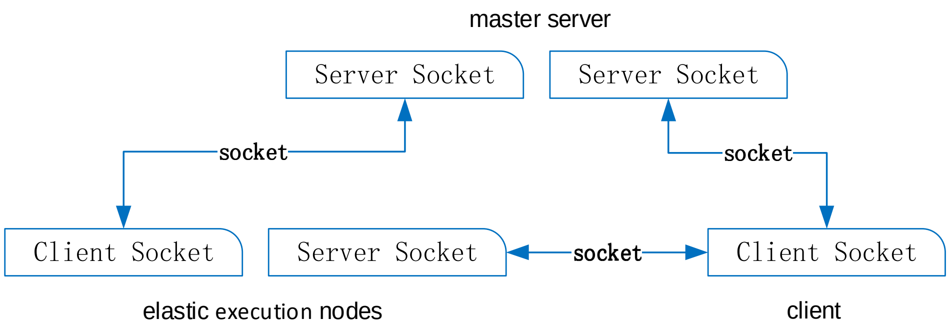


Figure 4-6. Communication architecture design of the server.

The link between the master server and elastic execution nodes is the link type between the master server as Server Socket and elastic execution nodes as Client Socket. The master server maintains the link information of all the current online elastic execution nodes and monitors the state of elastic execution node server. Elastic execution nodes are taken as Server Socket, and the client is taken as Client Socket. After the client acquires the node list from the master server, it links with the corresponding nodes and submits the water quality simulation task or scheme file download task. The master server runs specific Server Socket and is responsible for receiving the links of the client and receiving data. When the client sends water quality computing, scheme file download and other requests to the server, the server shall firstly add the client as a temporary link in the use list. Then, it reads the request content of the client and selects one or more data nodes to return to the client. After the client gets the feedback information, it cuts off the link with the server.

4.3.2 Method of acquiring the elastic node state by the master server

When the server stores the simulation scheme of the water quality model, the master server selects the node with specified number according to the principle of maximum current disk available space and least number of storage files. In the water quality simulation computing, the master server needs to know which elastic nodes own the strongest current processing capacity. Therefore, in the data server socket and scheduling server socket of the master server, each link needs to store the link state. In IOCP mode, each link of the server socket all owns an overlapping data structure. This structure can be defined according to requirements. To real-time monitor the state of the model computing nodes, byte array with certain length is set in advance in the overlapping data structure to save the state information of nodes, including CPU state, available memory, available disk space, average inflow, average outflow and maximum response time.

After elastic nodes are linked to the server socket of the master server, the timer is activated to regularly send the basic information to the server in form of data package. After the server receives the information, it updates the state of nodes in the overlapping structure.

4.4 The distributed file system of the water quality model

The distributed file system refers to the file management system where physical storage resources are not always directly linked to the local nodes and instead they are linked to nodes by the computer network (Thekkath, 1997). The distributed file system is one of the core technology of cloud computing as well as a key infrastructure of the distributed computation in the water quality model.

4.4.1 The current distributed file system

People have studied the distributed file system since 1970s. Not until network file system (NFS) in1980s did the distributed file system start to be applied in kinds of field. In the early stage, the distributed file system was represented by NFS and Andrew File System (AFS). By providing standard interface, remote file access was realized. With cloud computing technology, distributed computing technology develops rapidly and the distributed file system becomes one of the core technology of cloud computing.

At present, there are numerous mainstream distributed file systems, including Google File System (GFS) of Alphabet, Distributed File System (DFS) of RedHat, General Parallel File System (GPFS) of IBM, Lustre of SUN, and Hadoop Distributed File System (HDFS) of Hadoop. This paper takes HDFS of open-source Hadoop as an example to illustrate the main features of the current distributed file system.

Hadoop is an open-source cloud system construction and HDFS of Hadoop is the realization of GFS's open-source edition. HDFS is designed to be suitable for the distributed file system of big-dataset application program and it can be deployed on numerous cheap servers. HDFS adopts master-slave structure and it can be accessed by multiple clients at the same time. Its composition includes a name node and multiple data nodes. Name node controls file storage metadata information of HDFS and data nodes are in charge of storage and management of data block. When the client requires data storage, HDFS classifies files into small data blocks with fixed size and saves them on one group of data nodes. To realize the reliable storage of files, HDFS designs the following mechanism (Zhang et al., 2009).

1) Redundant storage. In HDFS, files are classified into small data blocks with fixed size and each data block is copied into many and saved on different data nodes to guarantee good fault tolerance of the file system.

2) Metadata disk failure. If the metadata of name node is destroyed and there is no data recovery mechanism, the whole the distributed file system will paralyze. HDFS adopts the backup strategy on the system metadata. Thus, the metadata can be recovered quickly after failure.

3) Cluster retrimming. To guarantee the storage balance of all data nodes in the distributed file system, when the disk space of one data node is occupied completely, part of its data block will be automatically quit by name node.

4) Error recovery. Data nodes cyclically send heartbeat packages to name node. Once data nodes fail, name node cannot receive the heartbeat packages of the node. The server will stop I/O operation of this node. When node failure is recovered and the replication factor of data block is lower than the specified value, name mode will copy these data blocks of this node.

5) Data integrity check. After the client reads data blocks in HDFS, data need to be verified and checked.

4.4.2 Requirements and architectural design of the distributed file system

(1) Requirements of water quality simulation on the distributed file system

under distributed architecture

To avoid the exchange of mass data between nodes, cloud computing adopts the migration strategy from computing to storage, which improves the execution efficiency of mass data processing. As the open-source implementation of GFS, HDFS has been widely and successfully used in distributed sorting, document clustering and statistics-based machine translation.

However, there is difficulty in directly applying HDFS to the water quality model. The major reasons are as follows. First, water quality simulation belongs to a computing-intensive process. By scheme file compression, the size of single-water quality simulation scheme file has been the secondary influence factor of water quality simulation performance in the mainstream network transmission environment, especially, in a LAN (Local Area Network) environment. The mechanism that the client directly transmits the whole scheme file and elastic nodes operate can be adopted. Second, in HDFS, files are split into many blocks and each is copied and saved on different data nodes, which cannot guarantee the complete storage of one file on a node. When the water quality model simulation operates, complete simulation scheme file sequence is needed. If the simulation scheme file is divided into different parts, and even a file is divided into many blocks and saved in different nodes, files need to be migrated and integrated to the computing nodes when the model is operated. It is inevitable to generate data flow direction contrary to mainstream cloud computing, namely, data flow from storage to computing, which reduces the distributed computing efficiency.

Therefore, the distributed computing of the water quality model needs to construct a proper distributed file system, where files are not further split. Instead, they can be completely saved on one or several elastic nodes.

(2) Architectural design of the distributed file system

In the operating system, software and data relevant to management documents are called file system. File system is an important part of the operating system and it usually lies in the bottom of the operating system. Realizing the file system at the bottom of the operating system can better combine with the operating system. However, its development difficulty and workload are huge. Considering these, the simulation scheme of the water quality model directly utilizes file programming interface provided by the operating system to store and retrieve data.

1) Storage design of the directory structure of the distributed file system

In the file system, directory and its subdirectory and files constitute the tree directory structure of the file system. Under the file interface programming design idea of the operating system, the directory structure in the distributed file system directly is a part of directory structure of the operating system. When elastic nodes are activated, working directory needs to be set, which is the physical directory of the operating system. When a new directory or file is created in the distributed file system, with the working directory as root directory, elastic nodes directly create it in the specified position according to directory tree structure.

2) Maintenance of directory structure of the distributed file system

In the master server, directory and file structure in the distributed file system are saved by the mode of the local database. The database contains two tables. One is the directory structure table of the distributed file system; the other one is the file storage node table of the distributed file system.

Table 4-2. Directory structure of the distributed file system.

Field	Type	Implication
ID	NVCHAR(36)	Unique identifier of distributed file directory or file
NAME	NVCHAR(30)	Mane of distributed file directory or file
PARENTID	NVCHAR(36)	distributed file directory or parent directory, and the parent directory of the root directory is null.

FSIZE	REAL	Size of the distributed file and the value of the directory is 0.
CTIME	DATETIME	The creation time of the distributed file
FTYPE	CHAR(1)	The value of the distributed file directory is noted as 0; the value of file is noted as 1.

Table 4-3. File storage node of the distributed file system.

Field	**Type**	**Implication**
ID	NVCHAR(36)	unique identifier
FID	NVCHAR(36)	unique identifier of the distributed file
NODE	NVCHAR(36)	unique identifier of elastic nodes saved in the distributed file

By the directory structure table of the distributed file system, the client acquires the directory structure. Inquiring the file list and storing the node list only need to inquire database by the master server. Files are stored in one or more elastic nodes. The client data can be downloaded from multiple nodes at the same time and further improve the file downloading capacity.

4.4.3 Scheme management analysis of water quality simulation

Based on water body grid partitioning in the research region, the water quality model simulation sets the basic parameters of the grid simulation, including initial water depth, initial flow and pollutant concentration of the observation station, water quality and hydrological data of time series, and completes the water quality simulation of the water body in the specified time period. Therefore, water quality simulation adopts scheme management. In the creation stage of the scheme, all the parameters of the simulation are set. Water quality simulation directly executes the executable file in the scheme to complete. Considering numerous simulation parameters in EFDC water quality simulation scheme, the water quality model parameters are stored in form of multi-file.

Each file of the water quality simulation is in form of document file and there is certain information redundancy. Because grid data are major data of the water quality simulation scheme, this paper takes grid data presentation and storage as an example to describe storage redundancy of EFDC water quality simulation scheme.

EFDC water quality simulation classifies the region of the river according to the same columns and rows, forming a surface grid system constituted of multiple quadrangles. In this grid system, each grid is a quadrangle and seamless connects with 8 adjacent quadrangles in space.

1) cell.inp file

This file stores cell location information and all grids are given integer variable to represent its type. For example, 5 represents water body, 0 represents land, and 9 represents land-water area. Figure 4-7 is a typical example of cell.inp. In this file, C labels behavior comment line. After the comment line, each line stores the line sequence of grid cells and the properties of each grid cell.

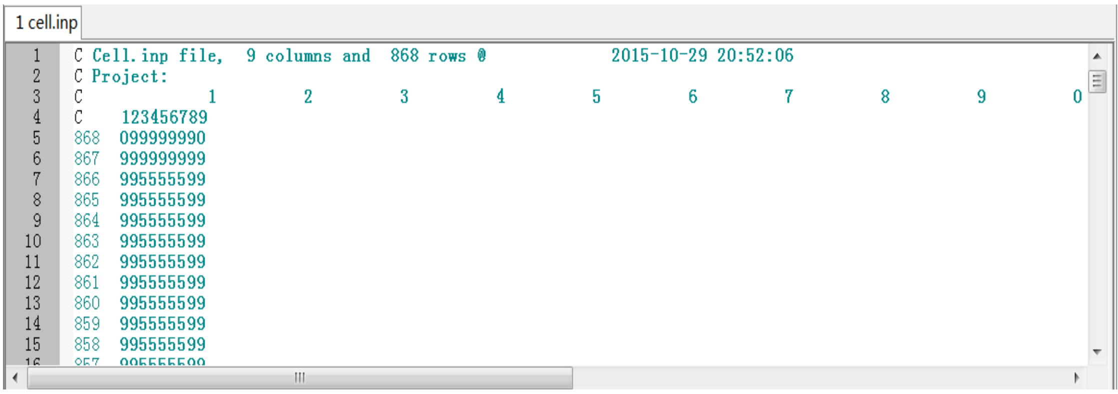
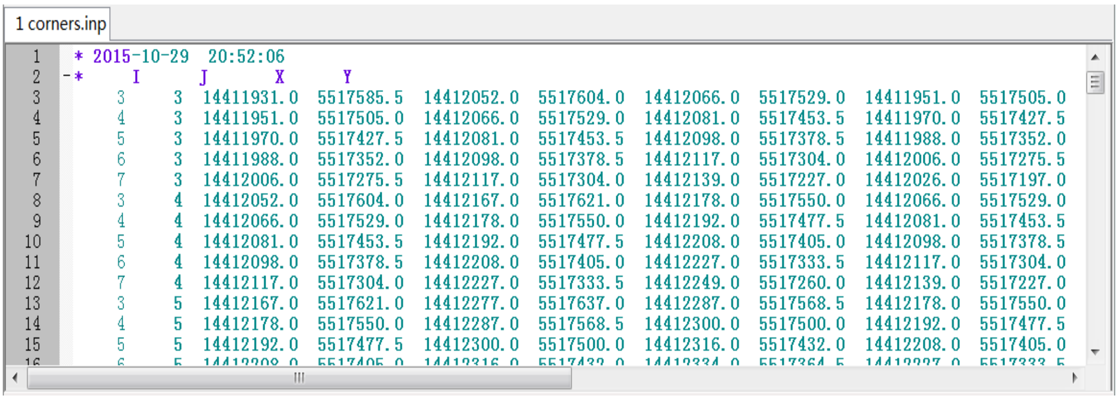

```
1 cell.inp
C Cell.inp file,  9 columns and  868 rows @           2015-10-29 20:52:06
C Project:
C          1         2         3         4         5         6         7         8         9         0
C    123456789
868  099999990
867  999999999
866  995555599
865  995555599
864  995555599
863  995555599
862  995555599
861  995555599
860  995555599
859  995555599
858  995555599
```

Figure 4-7. Example of cell.inp file.

2) corners.inp file

corners.inp stores the coordinate data of grid cells. In EFDC, each grid is a quadrangle. Figure 4-8 is a typical coordinate data file of grid cells. This file starts from the 3rd line, recording four angular points of each grid cell of water body.



1 corners.inp

```
* 2015-10-29  20:52:06
-*     I     J       X       Y
      3     3  14411931.0  5517585.5  14412052.0  5517604.0  14412066.0  5517529.0  14411951.0  5517505.0
      4     3  14411951.0  5517505.0  14412066.0  5517529.0  14412081.0  5517453.5  14411970.0  5517427.5
      5     3  14411970.0  5517427.5  14412081.0  5517453.5  14412098.0  5517378.5  14411988.0  5517352.0
      6     3  14411988.0  5517352.0  14412098.0  5517378.5  14412117.0  5517304.0  14412006.0  5517275.5
      7     3  14412006.0  5517275.5  14412117.0  5517304.0  14412139.0  5517227.0  14412026.0  5517197.0
      3     4  14412052.0  5517604.0  14412167.0  5517621.0  14412178.0  5517550.0  14412066.0  5517529.0
      4     4  14412066.0  5517529.0  14412178.0  5517550.0  14412192.0  5517477.5  14412081.0  5517453.5
      5     4  14412081.0  5517453.5  14412192.0  5517477.5  14412208.0  5517405.0  14412098.0  5517378.5
      6     4  14412098.0  5517378.5  14412208.0  5517405.0  14412227.0  5517333.5  14412117.0  5517304.0
      7     4  14412117.0  5517304.0  14412227.0  5517333.5  14412249.0  5517260.0  14412139.0  5517227.0
      3     5  14412167.0  5517621.0  14412277.0  5517637.0  14412287.0  5517568.5  14412178.0  5517550.0
      4     5  14412178.0  5517550.0  14412287.0  5517568.5  14412300.0  5517500.0  14412192.0  5517477.5
      5     5  14412192.0  5517477.5  14412300.0  5517500.0  14412316.0  5517432.0  14412208.0  5517405.0
```

Figure 4- 8. Example of corners.inp file.

In EFDC grid system, each grid cell is seamless connected to adjacent quadrangle space, namely, adjacent grid cells share one or two angular points. Coordinates of grid cells are separately stored and expressed. Shared angular points are repeatedly stored and there is certain redundancy.

3) celllt.inp file

The file content of celllt.inp file is completely consistent with cell.inp, but EFDC computing cannot lack this file.

4) dxdy.inp file

dxdy.inp file stores cell grid spacing, water depth, reservoir bottom elevation, reservoir bottom roughness and vegetation form information. In Figure 4-9, I and J represent codes of cells; DX and DY represent the horizontal and vertical spacing of cells. As properties of cells, DEPTH, ELEV, ZROUGH and TYPE respectively represent water depth, reservoir bottom elevation, and reservoir bottom roughness.

```
C DXDY.INP FILE, IN FREE FORMAT ACROSS COLUMNS for  4208 Active Cells
C 2015-10-29  20:52:06
C                                          BOTTOM                    Veg
C  I    J          DX        DY     DEPTH    ELEV     ZROUGH      TYPE
    3    3         79.5     119.9    0.700    245.65        0.035
    3    4         74.0     115.1    0.700    245.50        0.035
    3    5         70.5     110.9    0.700    245.34        0.035
    3    6         67.5     111.4    0.700    245.19        0.035
    3    7         64.1     101.2    0.700    245.03        0.035
    3    8         61.0      98.1    0.700    244.87        0.035
    3    9         58.8     100.0    0.700    244.72        0.035
    3   10         56.9     104.0    0.700    244.56        0.035
    3   11         55.1     105.3    0.700    244.45        0.035
    3   12         53.6     108.2    0.700    244.34        0.035
    3   13         52.5     106.0    0.700    244.23        0.035
```

Figure 4- 9. Example of dxdy.inp file.

Suppose four angular points of cells respectively are (X1, Y1), (X1, Y1), (X1, Y1) and (X3, Y3), then DX and DY can be calculated according to Formula 4-1 and Formula 4-2.

$$DX = \sqrt{[0.5 \times (X4 - X3 + X1 - X2)]^2 + [0.5 \times (Y4 - Y3 + Y1 - Y2)]^2} \tag{4-1}$$

$$DY = \sqrt{[0.5 \times (X4 - X1 + X3 - X2)]^2 + [0.5 \times (Y4 - Y1 + Y3 - Y2)]^2} \tag{4-2}$$

According to Formula 4-1 and Formula 4-2, corners.inp file can automatically generate by dxdy.inp file.

5) lxly.inp file

lxly.inp file stores center coordinate and rotation matrix of cells. Figure 4-10 shows the example of lxly.inp file, where I and J represent the codes of cells; X and Y represent the center coordinates of cells; CUE, CVE, CUN and CVN represent the rotation matrixes of cells.

```
C LXLY.INP FILE, IN FREE FORMAT ACROSS LINE for  4208 Active Cells
C 2015-10-29  20:52:06
C                                                                    WIND
C     I     J        X          Y        CUE       CVE       CUN       CVN   SHELTER
      3     3  14412000.0  5517555.9  0.17778  0.98407  -0.98407  0.17778  1.00  0.000E+00
      3     4  14412115.8  5517576.0  0.16529  0.98624  -0.98624  0.16529  1.00  0.000E+00
      3     5  14412227.3  5517594.1  0.15565  0.98781  -0.98781  0.15565  1.00  0.000E+00
      3     6  14412337.0  5517611.6  0.15926  0.98724  -0.98724  0.15926  1.00  0.000E+00
      3     7  14412442.0  5517628.3  0.15329  0.98818  -0.98818  0.15329  1.00  0.000E+00
      3     8  14412540.5  5517643.3  0.14781  0.98902  -0.98902  0.14781  1.00  0.000E+00
      3     9  14412638.5  5517657.6  0.14246  0.98980  -0.98980  0.14246  1.00  0.000E+00
      3    10  14412739.5  5517671.8  0.13477  0.99088  -0.99088  0.13477  1.00  0.000E+00
      3    11  14412843.3  5517685.1  0.12108  0.99264  -0.99264  0.12108  1.00  0.000E+00
      3    12  14412949.3  5517697.6  0.11325  0.99357  -0.99357  0.11325  1.00  0.000E+00
      3    13  14413055.8  5517708.6  0.09204  0.99576  -0.99576  0.09204  1.00  0.000E+00
```

Figure 4-10. Example of lxly.inp file.

Center coordinates of cells can be obtained according to Formula 4-3 and Formula 4-4.

X=（X1+X2+X3+X4）/4 (4-3)

Y=（Y1+Y2+Y3+Y4）/4 (4-4)

CUE, *CVE*, *CUN*, and *CVN* can be obtained according to Formula 4-5 and Formula 4-8. Similar to dxdy.inp, lxly.inp file can automatically generate by dxdy.inp file.

CUE=(0.5×(X4-X3+X1-X2))/DX (4-5)

CVE=(0.5×(X4-X1+X3-X2))/DY (4-6)

CUN=(0.5×(Y4-Y3+Y1-Y2))/DX (4-7)

CVN=(0.5×(Y4-Y1+Y3-Y2))/DY (4-8)

In conclusion, EFDC file redundancy mainly comes from two aspects, namely, storage mode and storage content. Storage mode refers to parameter files of EFDC in form of text file. Storage content refers to the information relativity between storage contents and part information can be deduced from other information.

4.4.4 Compression storage of water quality simulation scheme

To improve the operating efficiency of EFDC, give full play to the computing capacity of node servers, and realize off-line operation of water quality simulation and other realistic demands, water quality simulation process needs to be transplanted to the server side to complete. On the other hand, water quality simulation scheme is involved in numerous parameters and parameter setting need to be completed at the client in an interactive mode. The roles of the client and servers determine the necessary existence in the network transmission of scheme files. Because traditional schemes are organized in form of multi-file and text mode and there is certain information redundancy, network transmission becomes an important constraint of improving interactive performance between the client and servers. Thus, the compression representation method of water quality simulation scheme is discussed according to the shortcomings of EFDC software.

Jia Peng et al., (2015) raised an XML-based EFDC-ML (EFDC Markup Language) to reconstruct EFDC input files so as to adapt to the integration development of water environment management information system. This method describes the necessary information of EFDC simulation scheme file in XML file formats for the convenience of scheduling the traditional distributed system.

However, for the distributed system, this storage structure is still too overstaffing, which is not beneficial to the distributed shared transmission of water quality simulation scheme. Below will discuss the storage compression of water quality simulation scheme from two aspects, storage method and storage content.

（1）Water quality simulation scheme of monofile and binary mode

To improve the interactive performance between the server cluster and client, monofile and binary mode are adopted to make store and present water quality simulation scheme so as to reduce the grid transmission quantity of the water quality model computing and improve the interaction capacity between the server and client.

Monofile refers to putting all the information of water quality simulation scheme in a unified file. Figure 4-11 shows the storage structure of monofile water quality simulation scheme. In this file, the first part stores the basic information of the simulation scheme, including scheme name, simulation start time, end time, model computing step length, output step length and pollutants decay coefficient. The second part is divided into grid coordinate data, columns and rows of storage grids, and the coordinate information of each grid point. The third part is about grid attribute data, storage grid category, grid elevation and water depth. The last part stores the coordinates of trunk stream and its tributaries, sewage draining exit, intake and outlet, the initial concentration and water quality hydrological time series data.

To further reduce the storage size of the simulation scheme, its simulation grid, pollutant initial concentration and water depth are stored in a swap file by binary mode according to the storage structure in Figure 4-11.

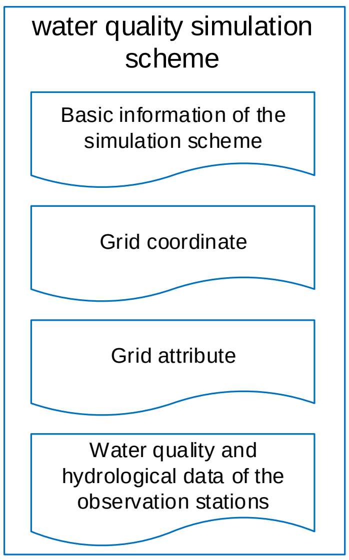


Figure 4-11. Storage structure of mono-file water quality simulation scheme.

（2）Storage content optimization of water quality simulation scheme

Shown as the above, water quality simulation scheme of EFDC software has numerous information redundancy. Taking grid data as an example, this paper elaborate the method of solving the redundancy of storage content.

By the analysis on the original EFDC software model files, storing the coordinates of storage grid cells can automatically generate the center coordinates and rotation matrix of grid cells by EFDC. Considering the sharing angular points of each grid and adjacent grids in the water quality simulation grid, though the coordinates of four angular points are respectively stored by taking grids as cells, there is still certain redundancy. Therefore, this paper takes the grid point as basis to express the grid data of the water quality model.

The grid points are described by the structure below.

```
struct GridPoint {double X; double Y;};
```

On this basis, the water quality model grid is described as,

```
class WaterMonitorGrid { GridPoint[] gCoords; int pointCols = 0; };
```

Where, gCoords represents the coordinate array of angular points in the simulation grids; and pointCols represents the total columns of the grid points. Suppose the row number and column number of the whole simulation grid respective are C and R, then pointCols is equal to C+1, and the array size of gCoords is $(C+1)*(R+1)$.

If binary stream is adopted for storage, point coordinates are stored in 16 bytes, and simulation grids are taken as cells to respectively store the four angular points of each grid cell, the total storage capacity S1 can be noted as,

$$S1=64*C*R \tag{4-9}$$

Taking grid angular point as cells to store the whole simulation grid, the total storage capacity S2 can be noted as,

$$S2=16*(C+1)*(R+1) \quad (4\text{-}10)$$

According to Formula 4-9 and Formula 4-10, without considering the binary system and text file storage difference, the storage capacity of the optimized grid coordinate decreases to about 1/4 of the original one. Because EFDC primary files are expressed in text file and several redundant files of cellt.inp, dxdy.inp, lxly.inp are included, the actual optimized storage capacity is smaller than 1/4.

To evaluate the information compression ability of water quality simulation scheme files in this paper, realistic simulation scheme files are made comparison test. Two grid sizes are set to respectively investigate the final storage sizes of different storage methods under the condition of storing the same information amount. To make the comparison structure more reliable, EFDC text file only investigates the total size of the replaced file sequence. The final comparison test is shown as Table 4-4. Compared with the original text file mode, the compression ratio of the binary file mode is close to 1/4, which can improve the network transmission performance between the client and servers to certain level.

Table 4-4. Comparison between EFDC text file storage and binary file storage.

Grid size	EFDC binary file (byte)	EFDC binary file (byte)	Compression ratio (%)
50*50	78, 342	292, 239	26. 8%
100*100	304, 494	1, 153, 519	26. 4%

4.4.5 Scheme file sharing and result storage of the water quality simulation

Under the distributed framework, water quality simulation includes two aspects, simulation scheme sharing and computing. For the sharing of water quality simulation scheme file, the client firstly submits the query request of the scheme list to the server. The server reads the scheme list of storage from the distributed file system and feedbacks to the client. The client selects to-be-downloaded scheme file and directly downloads them from the server.

For water quality simulation computing, the client firstly uploads local scheme files to elastic nodes. After elastic nodes receive scheme binary files, they are restored to EFDC water quality simulation text files. Then, executable files of the water quality simulation are called for water quality simulation. After elastic nodes complete the simulation, the client downloads the simulation results and reads them to GIS grid file format and displays them.

The distributed frame of the sharing of the water quality simulation scheme is shown as Figure 4-12.

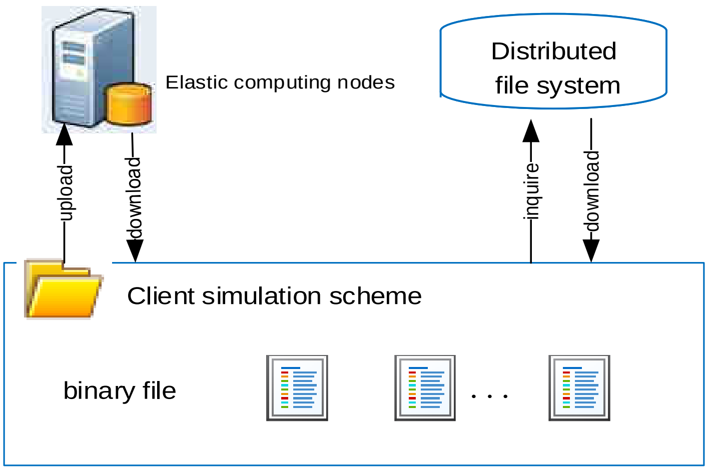


Figure 4-12. Sharing of the water quality simulation scheme.

The water quality simulation scheme initially is defined and generated by the client. Then, it is uploaded the distributed file system of the server for inquiry and downloading of all clients. To improve the file downloading capacity of clients, the simulation scheme is saved in one or more elastic nodes.

When the simulation scheme files are uploaded, the client firstly submits the inquiry request of candidate elastic nodes to the master servers. After the master server receives the request, elastic nodes are sorted according to the node storage balancing principle and elastic nodes of specified number are returned to the client. These nodes will be used to store the full backup of the simulation scheme. After the client receives the task list, local scheme files are divided into multiple subblocks and they are uploaded to all the candidate nodes according to block. After elastic nodes receive data blocks, whether simulation files exist shall be judged first. If no, a simulation file is created. Then, subblock is saved to the corresponding position. When all the sublocks on the client are uploaded, the simulation scheme file names and stored elastic node lists are uploaded to the master server, which registers the simulation scheme in the distributed file system. The uploading flow of Mudan River water quality simulation scheme file is shown as Figure 4-13.

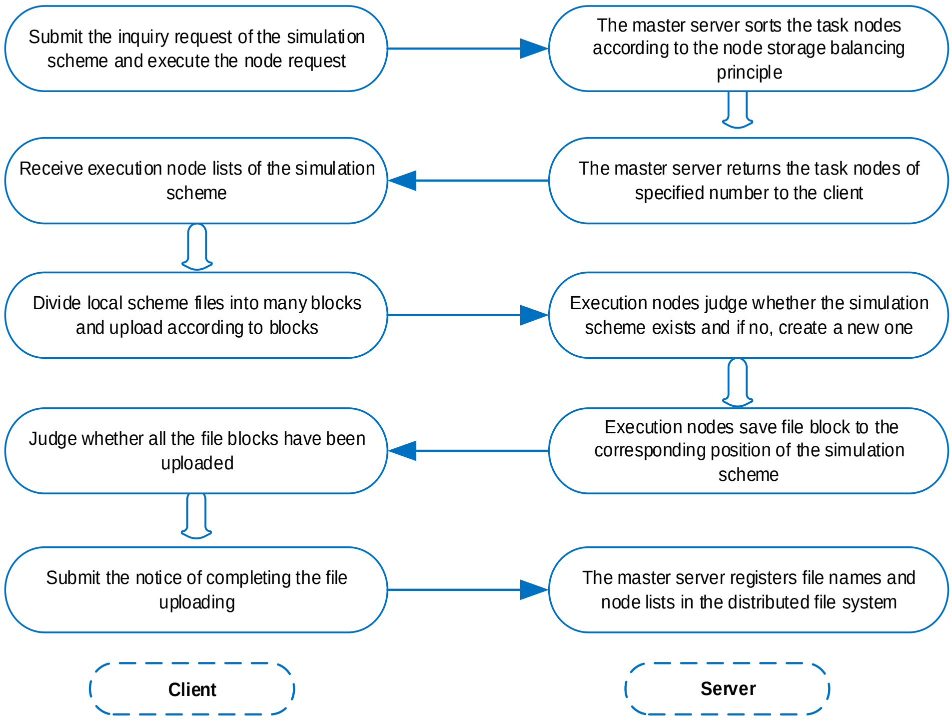


Figure 4-13. Uploading flow of Mudan River water quality simulation scheme.

After water quality simulation scheme is uploaded to the server, the client can inquire the scheme file and download the simulation scheme from the stored elastic nodes. Because the simulation scheme is stored on multiple elastic nodes, each node is in charge of downloading part data for the client, which guarantees the downloading speed of the simulation scheme.

The downloading flow of the simulation scheme by the client is as follows. The client firstly submits the downloading request of the simulation scheme. This request contains the full path of the target simulation scheme. The master server reads the stored elastic node lists in the distributed file system and returns them to the client. After the client receives the elastic node lists, it activates the load balance scheduling strategy to dynamically sort the elastic nodes and successively delivers downloading request of file subblocks to the optimal node. After elastic nodes receive the downloading request, they read subblocks from the local file and return them to the client. When the client completes downloading all the subblocks, the complete data of the simulation scheme are obtained. The downloading flow of Mudan River water quality simulation scheme file is shown as Figure 4-14.

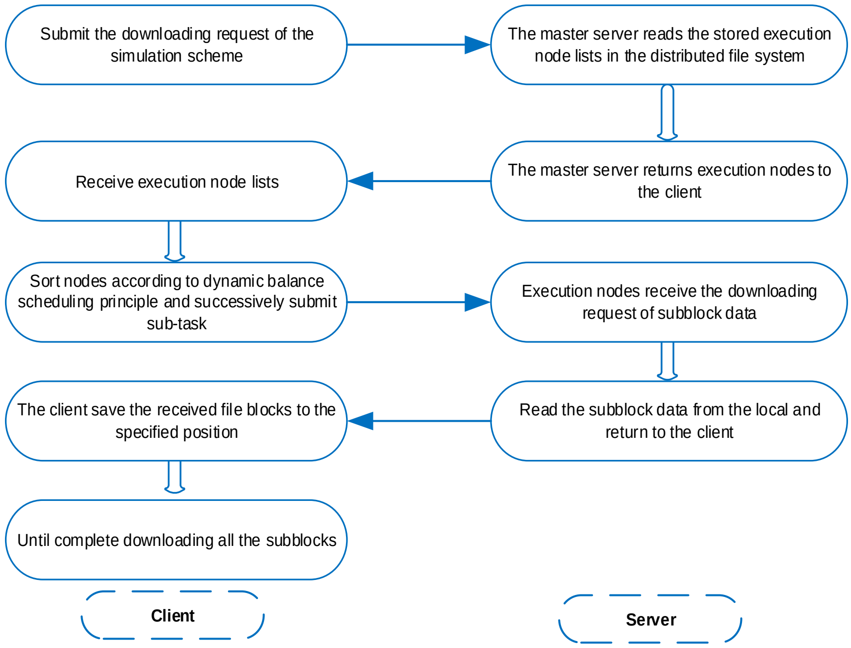


Figure 4-14. Downloading flow of Mudan River water quality simulation scheme.

As previously mentioned, to make the most of the computing resources of servers and realize the off-line operation of water quality simulation, the water quality simulation process is completed by the server cluster. Because the water quality simulation process is a computing-intensive process and cannot be subdivided, the water quality simulation scheme can only be completed by single elastic node. After the executable program of the water quality simulation is completed, simulation result files will generate automatically. The results are registered in the distributed file system, where the client can directly inquire whether the files of the simulation scheme exist and further whether the execution of the scheme is completed can be judged. If the execution of the scheme is completed, the client can visit the storage position of the simulation scheme in the distributed file system and download the simulation results. GIS functional interface of the client can read the downloaded simulation results. Visualized and categorized color rendering is made on the simulation results and exceeding water quality regions are showed with early-warning color.

4.5 Chapter conclusion

This chapter analyzed the integration method of water quality model of Mudan River and GIS in early warning system as well as the distributed architecture of dynamic elastic computing of water quality model. According to the calculation characteristics of water quality mechanism model, tightly coupled integration method was applied to this study. Considering the water quality spatio-temporal simulation and actual efficiency requirements of the water pollution emergency, we designed a distributed model computing method, distributed of the model calculation into the elastic node server in the cluster, made clear the dynamic elastic computing distributed architecture of water quality model. The communication method among server nodes of distributed water quality SEWS as well as the communication mechanism between servers and clients are designed under such architecture. In addition, considering the transmission problem of water quality model scheme files and combination of existing distributed file system, we designed the binary storage management methods of water quality simulation scheme file, design the distributed sharing way of model simulation program, file sharing and storage method of the model simulation results. This approach can significantly simplify data transmission among server nodes so as to improve the efficiency of distributed model computation under the same network condition. The distributed server’s elastic node and distributed file sharing are important basis to achieve the distributed computation of water quality model.

Chapter 5

Distributed computing and load balance scheduling method of water quality model

5.1 Connotation of distributed computing of water quality modeling

In the modeling schemes for daily and occasional prediction of water quality, pollution concentration and the relevant hydrological data of the monitoring stations need to be determined. These data can be obtained from the database. As water quality simulation is presented through schemes, the water quality and hydrological data are solidified in the modeling schemes. Thus, water quality modeling can be divided into two stages: scheme creation and water quality computation.

In the stage of establishing the water quality modeling scheme, the GIS spatial query would be used to obtain the initial pollution concentration and relevant hydrological data from the database. Flexible executing nodes are the specific executor of database query. In establishing the modeling scheme, the client packs the query water quality data and hydrological data in an inquire statement, and submits the query request to the master server. The master server selects the optimal node and returns it to the client according to the strategy to schedule the dynamic load balance. The client passes specific inquire statement to specific elastic nodes, and the elastic nodes complete data query and return the query results to the client.

In the stage of water quality computation, flexible executing nodes on the master server would be used to conduct the computation. Master server still function as an actor to schedule the tasks. Since water quality modeling is a process with intensive computation, when optimal node is selected, the number of simulation tasks under execution in the elastic nodes is the only criteria to select the optimal node. When the optimal node is returned by the master server, the client uploads the binary schemes of water quality simulation. Consequently, the elastic nodes interpret the binary file to generate a text file sequence which can be identified by the executable program of the water quality model. The calculation results are returned to the client in binary format and the client interprets the returned results and visualizes them as a new layer on the geographical map.

In the stage of water quality computing, simulation schemes can be stored in a distributed file system for sharing. Multiple elastic nodes are involved in upload and download of the simulation scheme, which is also a major demonstration of distributed computing.

5.2 Programming model of distributed computing

Map/Reduce is a kind of software architecture proposed by Alphabet for parallel computing of large-scale data sets (greater than 1 TB). It is through Map/Reduce, Alphabet processes above 1 billion queries from various countries every day, in 0.25 s on average (YANG, et al. 2007). Map/Reduce hides paralleling, load balancing, fault tolerance and other technical details, and provides convenient solutions for data processing of parallel system. As a kind of parallel computing technology, Map/Reduce has large distributed data-handling ability, and is one of the key technologies of cloud computing.

Before Map/Reduce was developed, programmers used various kinds of methods to handle mass of raw data, such as document fetching and Web request log. These data processing operations are easy to be understood in concept; however, because of the great size of data input, to complete the operation in an acceptable period of time, the computing must be distributed to a large number of node servers, while parallel computing, data distribution, error processing need abundant codes, making the whole process difficult.

In order to solve above-mentioned complex problems, engineering designers put forward an abstract process model. This model only requires stating the simple computations to be performed, instead of parallel computing, data distribution, fault tolerance and load balancing, etc., that are packaged in a library. The idea of developing this abstract model was derived from LISP and Map/Reduce primitive of some other functional language, i.e. Map/Reduce, the core technology of cloud computing (Afrati et al., 2010; Wu Baogui et al. 2007).

Map/Reduce refers to mapping and reducing. As shown in Figure 5-1, computing tasks are split into multiple independent but parallel subtasks. These subtasks are transmitted through network to elastic computing nodes to perform Map operation; Map operation input parameters are abstracted as a key/value pair, the results of operation generate key-value pairs of intermediate state, and a Reduce operation merges the results of the same key value, to get the final result set (Berthold, et al. 2009; Jeffrey, et al. 2008).

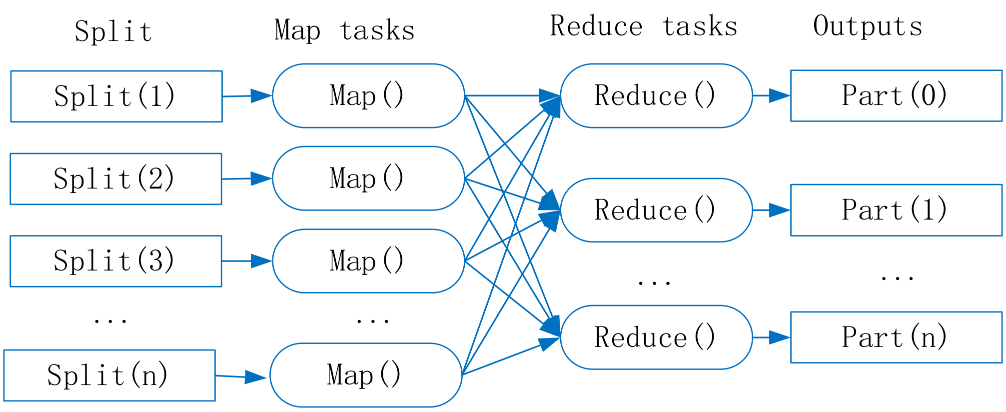


Figure 5-1. Programming model of Map/Reduce.

5.3 Task scheduling and load balance of distributed computing

In the framework of distributed computing, elastic nodes are actually executor of calculation. In distributed system operation, multiple clients simultaneously submit computing tasks to the server, and a task is subdivided into multiple subtasks, so that selection of elastic nodes becomes the key to distributed computing. If the task scheduling mechanism is unreasonable, a node or some nodes are overburdened, while other nodes are free, performance of the server will decline.

To achieve load balance, an appropriate node evaluation mechanism is required, firstly, to develop node evaluation methods that can accurately reflect the execution of nodes; secondly, to analyze state of the node state transmission. Appropriate task scheduling should timely and accurately reflect instantaneous state of notes, and make load balancing more reliable.

5.3.1 Main evaluation indicators of elastic nodes

Evaluation of elastic nodes is based on expression of the node state. Optional indicators of node state include available storage of nodes, CPU occupation, network transmission load, file storage and available memory, etc. At present, weight evaluation method is mainly used in evaluation, and different weights are set for different indicators, to indicate importance of the evaluation indicators.

Importance of indicators differs in executing tasks of different types. When a data storage command is submitted to the elastic node, storage capacity of nodes should be evaluated, then file storage, available storage space and network transmission load are more important than other indicators. When a calculation command is submitted to the elastic node, CPU occupation and available memory are more important. Therefore, balance of evaluation indicator is very difficult, and it should be analyzed by taking features of tasks into account. In this research, based on the characteristics of water quality model, the tasks were divided into two categories, data storage and data calculation, to discuss the load balance scheduling, respectively.

5.3.2 Load balance of data storage

Load balance mainly refers the selection of nodes for data storage when client file is uploaded to the server, to ensure storage balance of all notes and avoid data concentrates on some nodes. Data storage, as an indicator that can be quantified easily, it can achieve load balance easily. In this work, file storage, available storage space and network transmission load are taken as the balance scheduling indicators, and priority of the indicators is shown in Table 5-1:

Table 5-1. Priority of the load balance indicators.

Main indicator	The second indicator	The third indicator
Available storage space	File storage	Network transmission load

After a storage command is received by master server, nodes with insufficient storage space are removed according to file size and available storage space of nodes; the remaining nodes are arranged in ascending order according to file storage, and nodes with smaller data storage are selected firstly, to ensure balance data storage in the nodes; When file storage is balance, network transmission load of node determines selection of the final elastic nodes.

5.3.3 Dynamic load balance scheduling of node calculation

Dynamic load balance scheduling of node calculation studies about how to select the optimal nodes in computation, and bring computing power of each elastic node of the server into full play.

(1) Maximum response time of the elastic nodes

Weight evaluation is often used to evaluate computing power of elastic nodes. The method of weight evaluation considers CPU occupation, available memory, network load capacity and sums these indicators by weight. As affected by complexity of model computing tasks, it is difficult to determine weights of the indicators; and the optimal weights change in tasks of different categories. Therefore, compared to the method of weight evaluation, incorporating some indicators into a comprehensive one is more practical.

In evaluating the computing power of nodes, when the available memory, CPU idleness, network idleness of a node are in the optimal state, highest score is obtained, so does the computing power, generally, when the time of executing subtask is the shortest, and the max response time of nodes is the minimum. As a result, the max response time can be taken as a comprehensive indicator to evaluate computing power of nodes. Time of starting and ending a subtask on the elastic node needs to be recorded; and time of ending after a subtask is completed and fed back through the network shall be recorded. The interval between the two records is the time of execution. A time queue of executing a subtask is established, and then the recent max response time in the queue will reflect instantaneous computing power of a node.

(2) Dynamic load balance scheduling and node status transmission of the client

Only when node status is accessible for the dispatcher without delay, the final scheduling scheme can be feasible. Node status transfer can be dived into two types: strategy initiated by the executing node and strategy initiated by the client.

Strategy initiated by the executing node refers to real-time updating and regular sending the max response time from the executing node to the master server. Recent max response time of each executing node is recorded on the master server. It is easy to initiate a strategy, but it is not that easy to realize real-time update. The master server only records approximate state of the elastic nodes, which is suitable for the selection of the optimal elastic node at submission of the first batch of subtasks in the early stage of task scheduling.

Strategy initiated by the client is a dynamic balance scheduling, suitable for load balancing scheduling of high precision. In task distribution, time of subtask execution is recorded and feedback time of each elastic node is updated. In the next task distribution, the optimal node is selected according to dynamic evaluation results of the client, rather than getting the node state from the master server.

To sum up, in computing task execution, load balance scheduling can be initiated by the elastic node or the client. At the beginning of task submission, a certain number of elastic node are taken as the target node of task execution according to max response time from master server, and subtasks of the first batch are sent in turn; when the subtasks are fed back from the elastic nodes, dynamic scheduling strategy of the client is used to timely update and rank the time of execution, and the optimal nodes are selected dynamically and tasks are submitted.

(3) Load balance scheduling based on processing time

For the client initiated strategies, time of task execution is collected dynamically according to the state of task execution of the elastic nodes. A legal evaluation mechanism needs to be built on the client, to realize balance client scheduling. Chen Jun et al. (2013) put forward the processing time based load balance scheduling. The schedule takes total time of execution of the elastic nodes as the basis of selecting the optimal nodes of sending. Main principle of the scheduling: the client establishes a total time variable of task execution on each elastic node, i.e. the total time of updating the elastic nodes after the subtasks are fed back; new subtasks are submitted to the elastic nodes with the minimum time. The scheduling strategy maximizes computing power of the elastic nodes, and realizes overall balance.

(4) Download of the simulation scheme and Map/Reduce scheduling on data queries

Load balance scheduling based on processing time is adopted for download of water quality simulation scheme and data queries. In the distributed file system, water quality simulation files are saved in different elastic nodes. In downloading the files, the download tasks are split in accordance with the Map/Reduce requirements, each subtask corresponds to a part of the scheme file. The number of subtasks of file download is calculated based on the split size, and these subtasks are submitted to the optimal elastic nodes, the nodes complete Map operation, read the sub-block file and download in the client; The client conducts Reduce operation against the files obtained and forms the final scheme. Scheduling process of downloading the water quality simulation scheme is shown in Figure 4-14.

In the process of download, the client request for file download is submitted to the master server, the master server searches the node list of data saved in the distributed file system, judges the availability of each node in the list by querying the real-time online node sequence saved in the master server, ranks the effective elastic nodes according to the average time of executing the minimum tasks, and returns the sequential node list to the client. The client splits the download tasks, and downloads unit data from a node. Unit data returned to the nodes are integrated on the client side, total time of execution of the elastic nodes is updated, and on this basis, elastic node with the shortest time of processing is taken as the candidate node to download the next unit data. After all the tasks are completed, the client gets the complete simulation scheme file.

In establishing the scheme of water quality simulation, water quality and hydrological information should be obtained from the database, and dynamic balance scheduling is still adopted. Nodes as actual executor of data query, read the query command from the client, get records from the database and feed back to the client in binary stream, and the client transfer data stream to get the results. Different with the scheme download, the operation of data query is indivisible. Each data query composes a unit task, Map operation is a query operation of data node, Reduce operation refers to the process of getting data set by the client. To realize overall balance in elastic node scheduling, execution time of the elastic node is taken as a global variable and stored in the client. When a new query task is submitted, historical data of the elastic nodes are read directly, and the optimal nodes are selected.

5.4 Distributed computing of water quality simulation

Dynamic balance scheduling based on processing time is a scheduling algorithm after event, i.e. current state of the node is evaluated after the current subtask is completed, which is suitable for the situation when a task is spilt into several subtasks with small time slice. In Mudan River water quality simulation, single water quality simulation is independently completed by an elastic node. Because of the complexity of water quality simulation, and as affected by long time of simulation and short interval of output, a single process of water quality simulation takes a long time. If the method of processing time based dynamic balance is still used, the node state cannot be fed back timely, and computing power of the elastic nodes cannot be evaluated accurately before the water quality simulation is completed, thus the scheduling method is invalid. Therefore, in view of the particularity of EFDC, a balance scheduling method based on the number of water quality simulation execution is presented in this study.

5.4.1 Water quality model computing thread pool of the elastic nodes

As water quality simulation is a compute-intensive process, powerful computing ability of the server can realize resources sharing with clients. The scale of elastic nodes is a major indicator on computing power of the server. Obviously, the more the elastic nodes cause the concurrent executing ability to become the stronger. In theory, there should be enough elastic nodes in the server for dynamic invocation, but in the actual cases, the number of elastic nodes may be limited. To enhance the adaptability of distributed computing, it is necessary to study the application of a single elastic node dealing with several simulation tasks. In the distributed computing framework, elastic nodes parallel computing requests by thread pool.

A certain number of thread pools are set according to actual computing power of the elastic nodes. For example, if M tasks of water quality simulation are executed on the elastic nodes, maximal size of the thread pool is M. Initial threads of the thread pool are marked as free, and when a computing task is received from the client, it will firstly check whether there is a free thread in the thread pool. The free thread will receive a task and become occupied; if there is no free thread, it will wait. After the thread completes the computing, it will be marked as free.

If the water quality simulation takes long time, the scheduling algorithm based on processing time fails, and the number of busy threads on elastic nodes becomes the only indicator of master server in selecting the optimal node.

5.4.2 Load balance Map/Reduce scheduling in water quality simulation

In this research, as a single simulation calculation is completed by at least one CPU, in water quality model computing, Map operation is the computing of elastic node; Reduce operation is the download and visualization of the simulation results in the client. In scheduling tasks of small time slice, maximum response time is generally used to evaluate computing power of the nodes. But in water quality model computing and scheduling, the number of concurrent execution of elastic node model becomes the only indicator for the optimal node selection. It is because water quality simulation is a compute-intensive process, and available computing power of the server is directly determined by the number of concurrent execution. As mentioned above, there is a thread pool in each elastic node, the number of busy threads in the pool is the number of concurrent execution in water quality model computing of the elastic node, and the value is packed to the master server together with CPU state, available memory, available disk space. When the master server receives the request of model computing, the o

The elastic nodes send node state to the master server on time, but state information sent to master server by elastic nodes is not timely. To ensure reliability of the elastic nodes finally selected, when the master server receives a water quality model computing request from the client, it selects some servers (candidate elastic nodes from 1 to M) based on the available information and send them to the client, and the client, after receiving the candidate elastic node list, make real-time contact with the elastic nodes, to get the latest information about the number of tasks on the nodes. On this basis, the nodes are ranked in ascending order, and the first one is generally the optimum.

To improve the reliability of task execution, the above method is not sufficient for selecting the optimal node. If the selected optimal one fails to carry out new task of water quality simulation as affected by the server (the target node fails to return the success of delivery within the allotted time or directly returns the information of failure), the remaining nodes will be ranked again, to find out the optimal elastic nodes and deliver again, till the tasks are delivered successfully or the elastic nodes in the list are used until exhausted.

Dynamic balance strategy of water quality simulation is as shown in Figure 5-2:

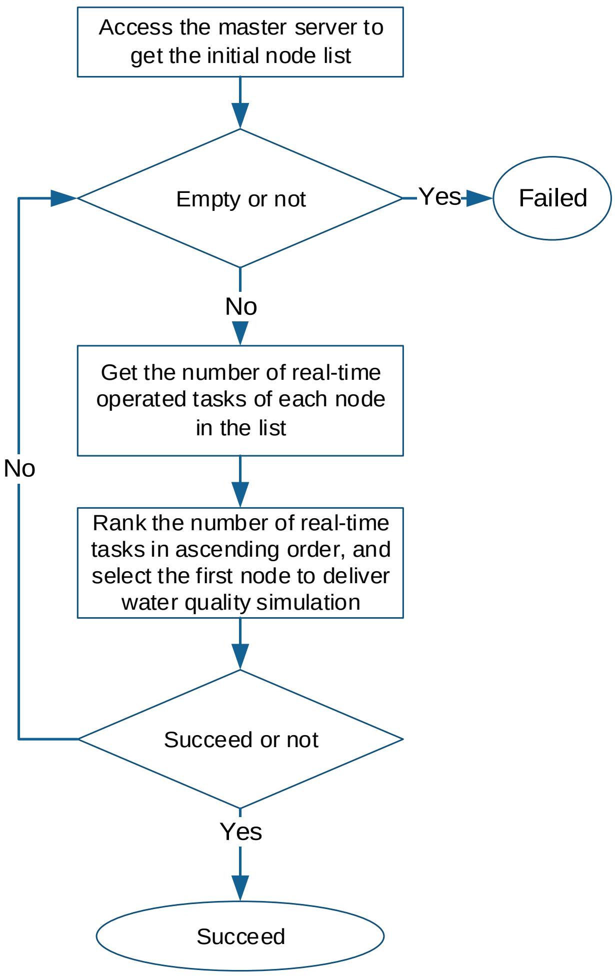


Figure 5-2. Dynamic balance strategy of water quality model computing.

5.5 Chapter conclusion

After the introduction to distributed programming model, this chapter discussed the task-scheduled load balancing of distributed computation, analyzed the main evaluation indexes of elastic nodes, including available storage of nodes, CPU occupation, network transmission load, file storage and available memory, etc. Next, this chapter discussed load balancing computation of data storage and dynamic load-balanced scheduling of elastic node computation and clarified Map/Reduce load-balanced scheduling of water quality simulation scheme. Furthermore, this chapter analyzed the distributed computation of water quality simulation and discussed the water quality simulated computation thread pool and Map/Reduce load-balanced scheduling strategy to solve the parallel computing problems of multi-model schemes of single node. The role of elastic node thread pool is used to improve the simulation capability of single server node scheme. Map operations of Map/Reduce load-balanced scheduling were simulation process performed by the elastic node, Reduce operations were the simulation results and presentation process downloaded by the client. To ensure reliability of the elastic nodes finally selected, when the master server receives a water quality model computing request from the client, it selects some servers based on the available information and send them to the client, so that the client could get the latest information about the number of tasks on the nodes. On this basis, the nodes are ranked in ascending order, and the first one is generally the optimum. In order to improve the reliability of the task execution, when choosing the best nodes, the failure information of task-performing node was processed by the method of cycle checking.

Chapter 6

Implementation of core technologies of Mudan River quality SEWS

6.1 Logical structure and functional framework of water quality SEWS

Water quality SEWS is logically divided into physical layer, data layer, application layer and presentation layer as shown in Figure 6-1. Physical layer is the bottom layer, mainly including such information infrastructure as computer system, storage system, network, safety management, and so on to support the system with physical resources.

Data layer is a very important component of water quality SEWS to provides data storage, data invocation and visit support for the system. Main content of the layer is a database platform (Oracle or Microsoft SQLSever) for organization and management of various data. The database includes basic geographic information data, business attribute data, pollution risk source data, environmental sensitive objects data, monitoring data, and so on.

Application layer is a software platform for professional technique services hence is used to solve the practical problems in service operation. It can be divided into 3 sub-layers: basic platform layer, basic application layer, and service application layer. Basic platform layer includes three major application software used in the system: GIS development functional component, EFDC software, and ArcGIS Desktop, which are respectively the software functional component, the professional environmental fluid dynamics model construction tool, and the GIS data editing and updating environment. Basic application layer includes spatial data inquiry, attribute information inquiry, water quality index adjustment, thematic information control, thematic data maintenance and system management, etc. Service application layer mainly includes the functions like water quality data evaluation, water quality trend analysis, water quality forecasting and early warning, daily and emergent water quality simulation, report forms and thematic map printing, etc. Presentation layer provides connector for relevant application program and man-machine interaction to implement the communications of logical processing results between the operator and the application server. Task request and scheduling submitted by the operator or the application system are also done through the layer.

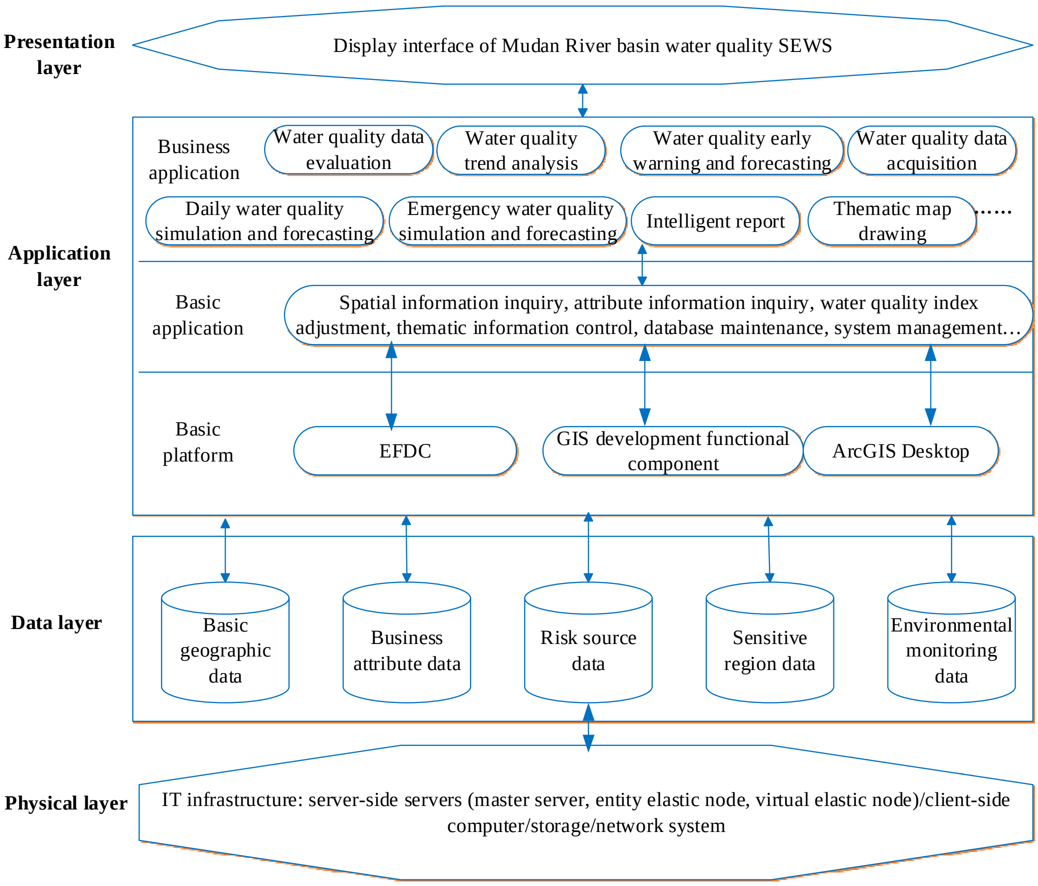


Figure 6- 1. Logical structure of the system.

As shown in Figure 6-2, main functional structure of Mudan River water quality SEWS mainly includes functions like system management, data acquisition, basic GIS functions, water quality evaluation and simulation & early warning, etc. The core technologies are database technology, water quality trend analysis, water quality evaluation, daily water quality and emergency water quality simulation and forecasting as well as standard-exceeding warning technology. Subsequent contents of this chapter will respectively discuss about implementing methods of these core technologies.

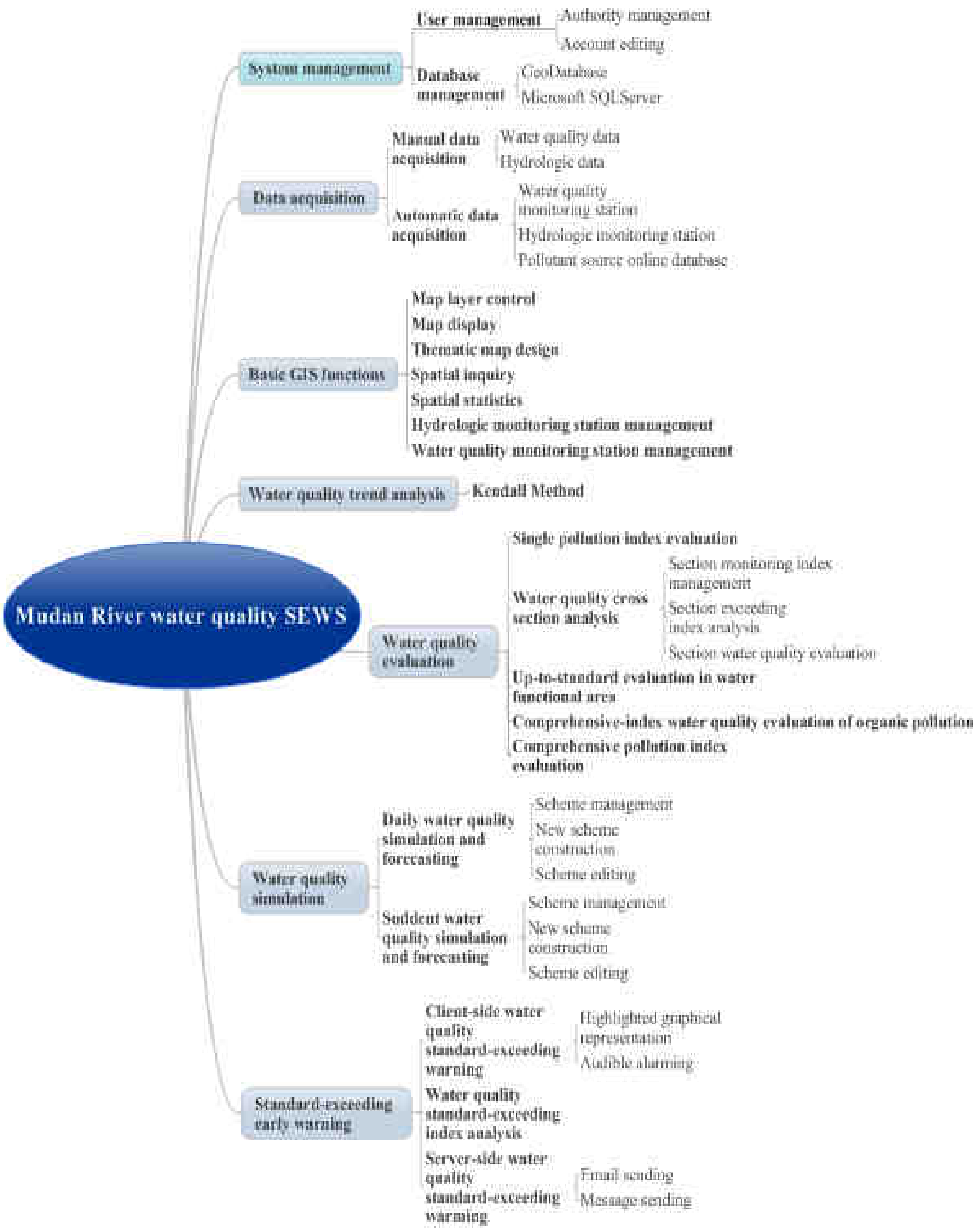


Figure 6-2. Schematic diagram of system functional structure.

6.2 Database of water quality SEWS

6.2.1 Spatial data processing method

Spatial database in this study contains data like watershed, lakes, rivers, reservoirs, digital elevation model (DEM), satellite images, etc. with many data varieties. Data variables may be mutually correlated or have strong correlation. Besides, it involves combined action of attribute data and spatial data as well as coupling and analytical processing of information of different scales, so this study adopts geodatabase to store relevant spatial data.

Geodatabase is a kind of intelligent spatial database structural model. It can conduct a uniform description of geospatial elements processed and expressed by GIS such as vector, grid, network and address under a uniform model framework, and in Geodatabase model, expression of geospatial elements is closer to cognition and expression mode of real objects, and it's a kind of data storage structure object- oriented. Geodatabase introduced behavior, rule and relation of geospatial elements. When elements in Geodatabase are being processed, it's not necessary to program coding of rules and basic behaviors which must be satisfied by these elements; for their special behaviors and rules, customization definition can be conducted through element expansion (Maguire et al., 2008), which is greatly beneficial for management of spatial data and their attribute data and program functional development based on this data model, Figure 6-3 is schematic diagram of Geodatabase.

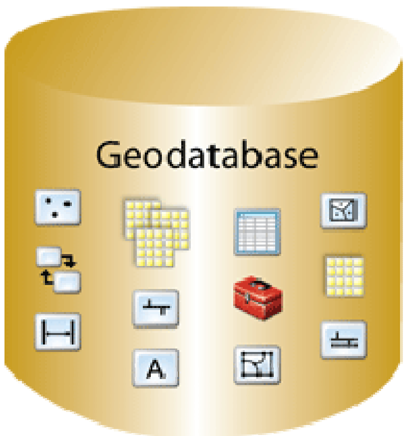


Figure 6-3. Geodatabase spatial storage model facing objects.

6.2.2 Relational database table

Business attribute data in Mudan River water quality SEWS may have spatial location information, but they are not stored in geodatabase such as precipitation in the basin, water quality monitoring, hydrologic monitoring, risk source and other service attribute data, and these data are stored in Microsoft SQL Server 2012 relational database platform in the form of table.

6.2.3 Basic contents of water quality early warning database construction

Basic contents include Mudan River basin geospatial data, hydrologic water quality monitoring point data, risk source data, pollution discharge declaration data, pollution accident case data and hydrodynamic water quality model parameter data, etc.

Geospatial data is data foundation for constructing GIS relevant system. In this study, it includes base map data like administrative region, traffic road, lake, digital elevation model and satellite image with uniform space coordinates information.

Hydrologic water quality monitoring point data includes river water quality monitoring point and hydrologic monitoring point information whereby the former mainly monitors all kinds of water quality parameter information while the latter can provide hydrodynamic model for usage in input parameters.

Risk source data mainly refers to risk-source enterprises nearby river and lake including pollution risk source enterprises and hazardous chemical enterprises. It generally includes monitoring indexes like COD, NH_3N, total phosphorous (TP), pH value, temperature and flow quantity. Hazardous articles data include name of hazardous article, quantity of hazardous articles, enterprise information, risk source location, information about industry involved, etc.

Pollution discharge declaration data includes declaration year, administrative division code, company name, legal representative, legal person code of declaring company, preparer, submit data, pollution discharge type, contact, tell/fax, e-mail, postal address, postal code, commissioning (business starting) date, production (management) days of the last year, original values of environmental protection facilities, number of sewage treatment facilities, disposing capacity of sewage treatment facilities, operational expenses of sewage treatment facilities of the last year, sewage discharge license number, date of issue of sewage discharge license, whether it's key pollution source, paid total amount of sewage discharge fees of the last year, environmental illegality penalty of the last year, etc.

Pollution accident case data: it stores relevant pollution accident cases and treatment strategies, it's a generalization and summarization of relevant experience knowledge and can provide valuable information for water environment emergency monitoring service. It usually includes basic conditions of accident case, dangerous objective, emergency rescue organization, personnel and responsibilities, alarming, communicating mode, process management measures that should be taken after the accident happens, personnel urgent evacuation, withdraw, detection, emergency rescue, rescue and control measures, emergency rescue guarantee, graded response conditions to plan, closing program of accident emergency rescue, emergency training plan, exercising plan, etc.

Hydrodynamic and water quality model parameter data provide hydrodynamic and water quality parameter information which are necessary for spatial-temporal variation simulation of pollutant concentration including grid, water depth and temperature, flow velocity, longitudinal dispersion coefficient of river section, bed roughness, pollutant attenuation velocity coefficient of river section, simulated spatial step length of the model, etc.

6.3 Implementation analysis of core technologies of water quality SEWS

6.3.1 Construction of water quality evaluation technology

River water quality evaluation is the groundwork for watershed environment monitoring management. Through adopting reasonable water quality evaluation technology, river water quality situation can be clearly illustrated and water environment management and decision-making requirements can be satisfied. Water quality evaluation module of SEWS establishes evaluation standard and selects evaluation parameters, evaluation method and spatial scope of evaluation mainly according to water quality evaluation objective and with a reference to relevant standards (e.g., GB3838-2002), and then it evaluates whether indexes of water quality monitoring stations within a certain spatial scope have reached water quality evaluation objective, and then it conducts a statistics of up-to-standard situation of monitoring stations within a certain spatial scope according to evaluation results and by classifying into administrative region and water functional region.

During implementation process of water quality evaluation functions, evaluation period and region, evaluation method and pollution factors shall be firstly considered; then concentration sequence of monitoring station pollution factors within special scope will be read from database according to these parameter information. Period selection has certain flexibility such as week, month, year, etc., we assume that t months, m stations and n pollution factors are selected, then there will be $t*m*n$ concentration sequences, and lengths between sequences may be different. Then we take standard values corresponding to n factors from database, according to surface water quality evaluation standard, each factor has 5 standard values which constitute a $5n$ matrix. This study takes the most unfavorable factor method as an example, and concrete implementation method of water quality evaluation technology is shown in Figure 6-4.

6.3.2 Construction of water quality trend analysis technology

Water quality trend analysis aims at mastering change rule of water quality with time (Hu et al., 2004). When water quality trend analysis is conducted, we can select month, season or year as time period and water quality monitoring values of different water quality monitoring sections to conduct the analysis, water quality trend analysis can be single-factor water quality evaluation and also can be water quality grading trend analysis. Taking seasonal Kendall method as an example in this research to analyze water quality trend.

Principle of seasonal Kendall inspection is comparing water quality data of the same months (or the same season) over the years, if posterior value is higher than anterior value, it will be recorded as "+", otherwise, it will be recorded as "-". If the number of "+"s is greater than that of "-"s, it may be rising trend; otherwise, it may be descending trend. If water quality data don't have rising or descending trend, then numbers of "+"s and "-"s respectively occupy 50%. River flow quantity is of yearly periodical changes, and concentrations of most river water quality components are influenced by periodic change of water environment, such as flow quantity, flow speed, temperature. Hence, it may be shortage of comparability of water quality data between ice-covered and open-water period. Seasonal Kendall inspection compares water quality data in the same month or the same season over the years, which can avoid seasonal influence. In the meantime, as data comparison only considers relative arrangement but not its size, it can avoid common missing value problem in water quality materials (Peng et al., 2014).

For seasonal Kendall method, time period, river and pollution factors of water quality trend analysis are firstly select, and pollution factors can be directly obtained from database. Then according to selected parameters, corresponding concentration sequences will be obtained from system database.

It’s assumed that there are m monitoring stations, *n* pollution factors, then there will be *m*n* concentration sequences, and lengths between sequences may be unequal. Secondly, it compares monitoring values of the same month (or the same season) of *m* monitoring stations within water quality trend analysis period. Then rising and descending variation trend will be obtained, or there will be no variation. Finally, the sum of values of rising trend, descending trend and no variation trend of water quality factors in monitoring stations within this water quality trend analysis period. Variation trend is divided into highly significant rising trend, significant rising trend, no-variation trend, significant descending trend and highly significant descending trend. Figure 6-5 displays technological implementation of water quality trend analysis with seasonal Kendall method.

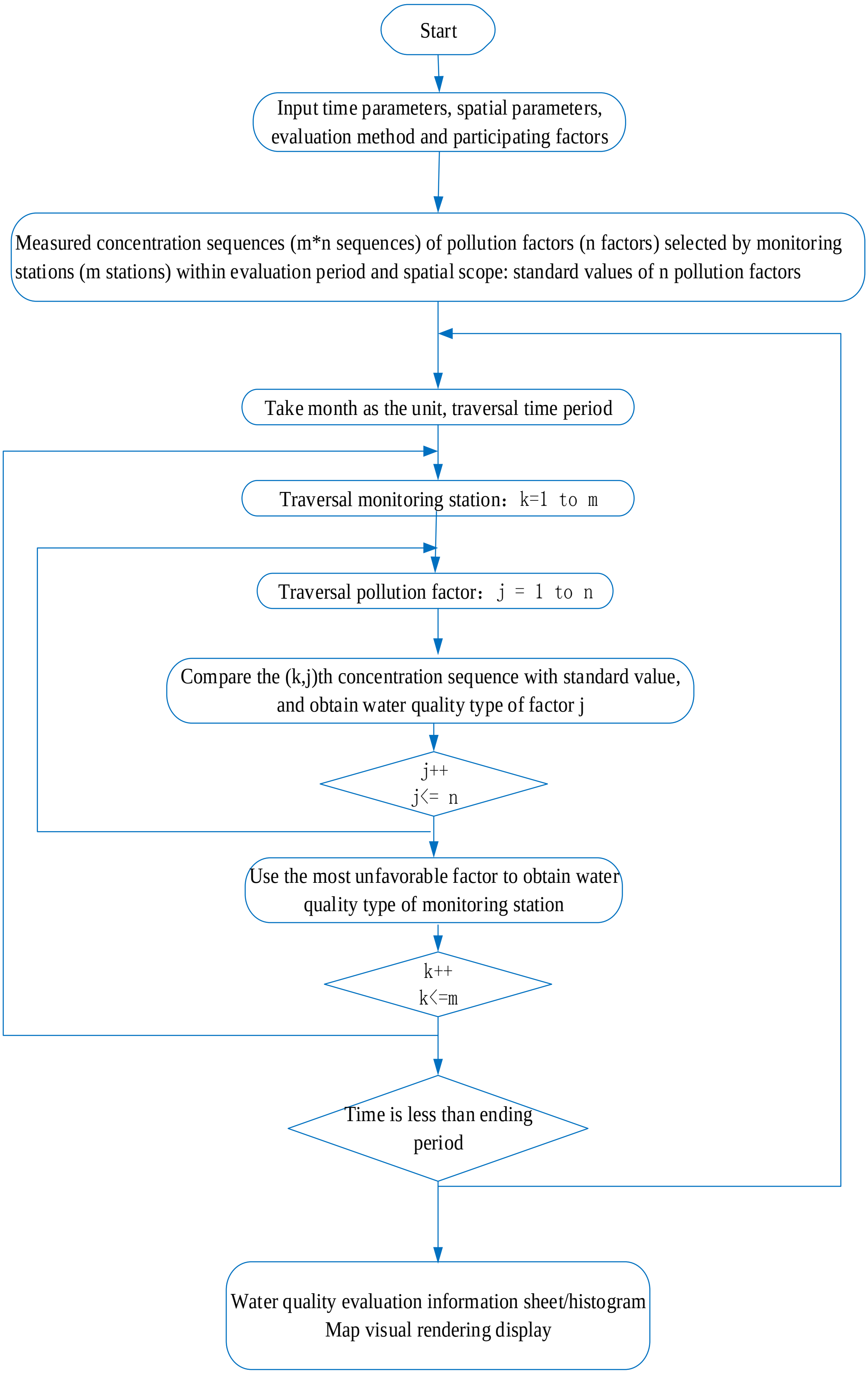


Figure 6-4. Water quality evaluation process.

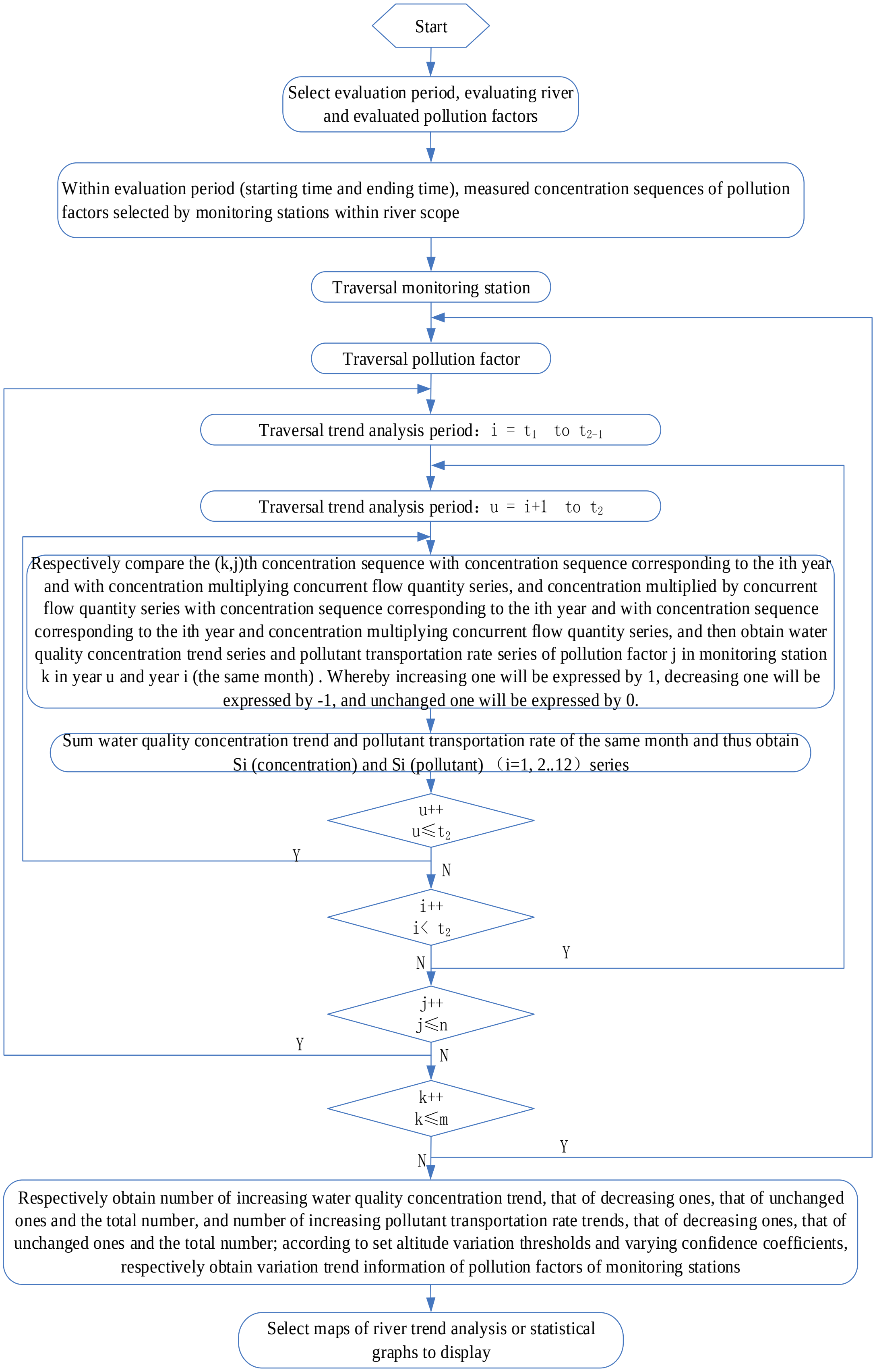


Figure 6-5. Water quality trend analysis process.

6.3.3 Construction and analysis of exceeding standard early warning technology

In early warning and forecasting process of water quality exceeding standard, early warning grade setting can adopt means of manual threshold setting or automatically read from relevant standards, and alarming is conducted by means of different colors rendering on the map, message, e-mail and so on according to functional requirements. The system conducts dynamic monitoring of factors set by monitoring stations, and provides automatic monitoring functions. It manages alarming information, time and grade.

According to characteristics of water quality exceeding standard alarming, implementation of this technology can be divided into alarming based on server-side mode and that based on client-side mode. The server periodically calculates water quality monitoring pollution indexes of database, automatically obtains exceeding standard monitoring stations and indexes, and then informs relevant personnel by means of e-mail. Auto early warning of the server is automatically finished on the server. The system timely starts according to exceeding standard trigger mechanism, makes intelligent interpretation of latest early warning time and about whether it’s necessary to update early warning, and technical route of server auto early warning is shown in Figure 6-6.

Different from early warning scheme of the server, client-side early warning technology can be artificially and repeatedly simulated and regulated. It can repeatedly conduct simulation and early warning analysis of the current pollution situation and display spatial-temporal dynamic simulation results of water quality. As the identification about whether e-mail has been sent or not has been recorded in the database, for the same time point, the person in charge of sending e-mails will display “e-mail sent” but will not repeatedly send. Technical route of client-side early warning is shown in Figure 6-7.

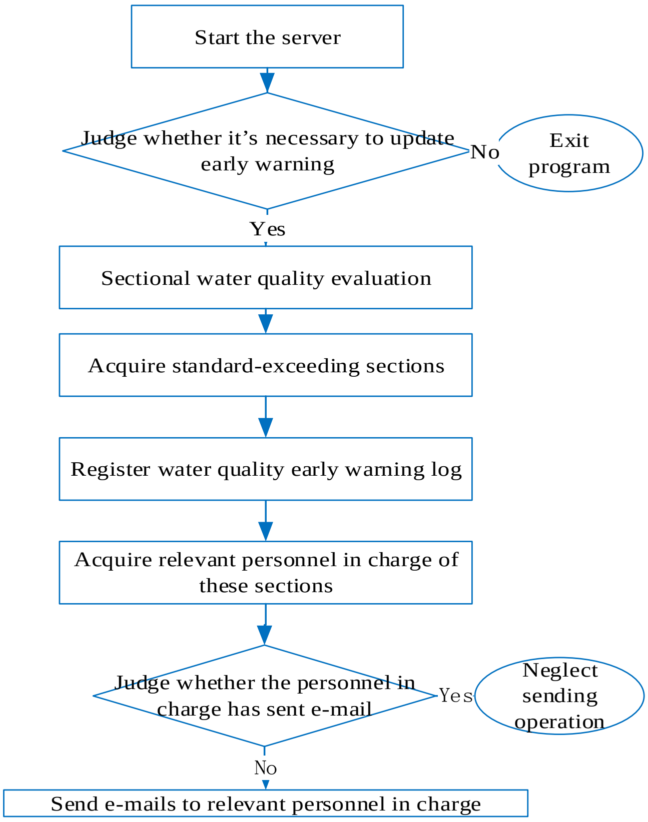


Figure 6-6. Technical route of server auto early warning.

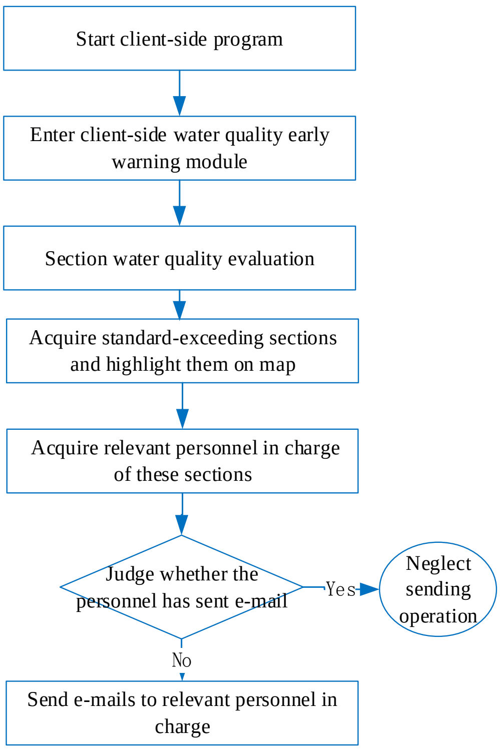


Figure 6-7. Technical route of client-side early warning.

6.3.4 Construction of conventional water quality simulation and early warning technology

Based on analysis of river water quality variation mechanism, water quality simulation and early warning uses hydrodynamic force-water quality coupling method to construct response relationship between water quality parameters and pollutant input which reflects change rule of water quality with pollutant input. Water quality change rule reflected by mechanism model is usually of universality, and model parameters should be calibrated and verified according to concrete river conditions. This study uses the model established in Chapter 3 to develop water quality simulation and forecasting module, uses established model to construct Mudan River water quality forecasting and simulation scheme, and the system provides functions like construction, management and analyze simulation results and so on of simulation scheme.

Construction and simulation process of Mudan River water quality model is shown in Figure 6-8, it’s necessary to firstly determine simulated river reach and starting & ending time and then divide grids; secondly, it’s necessary to further determine scheme of boundary conditions, and input flow quantity of entrance, exit, intake and branch import of simulated river reach and background concentration of pollutant; then it needs to determine scenario scheme of pollution load discharge and then input flow quantity discharged at drainage exit and pollution load capacity; it determines the background concentration of pollutants in each cell of 2D grids; according to preset scheme, it determines used model parameters and generates initial data into binary files needed in the simulation and uploads to the server. Selection and operation of server as well as uploading and downloading process of binary files are presented in Chapter 4; finally, it reads water quality simulation results information, visual water quality status and spatial-temporal propagation process of dynamic simulation from output files.

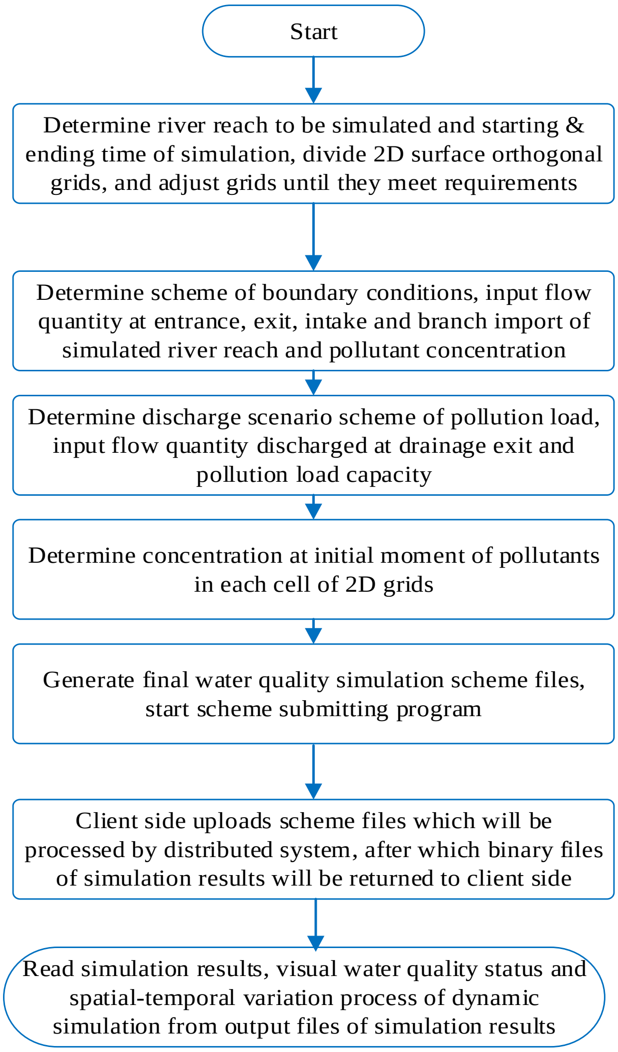


Figure 6-8. Construction and simulation implementation of mudan river water quality model.

6.3.5 Construction of emergency water pollution simulation technology

Implementation process of emergency water pollution early warning technology is largely the same with that of conventional water quality monitoring and early warning, the program needs to take consideration of occurrence site of sudden pollution accident and implement calibration of occurrence coordinates of pollution accident by virtue of interaction map, and the system will automatically determine its position in gridding system of Mudan River model according to this position. Verified Mudan River water quality model has been constructed in Chapter 3, emergency water pollution early warning can use parameters like grids, bed roughness and pollutant attenuation rate determined by this model to conduct the simulation. According to simulation results and under the support of GIS system, it can conduct spatial-temporal visual variation process analysis of pollutants for the pollution accident which has been simulated.

To implement emergency water pollution simulation technology, it's necessary to determine starting & ending time of pollution accident, the river where it occurs, pollution type and pollution factors, then according to these information, it will search for corresponding pollution accident scheme template from database. If existing pollution accident scheme can't meet requirements, it's necessary to construct a new simulation scheme, and concrete flowchart is shown in Figure 6-9.

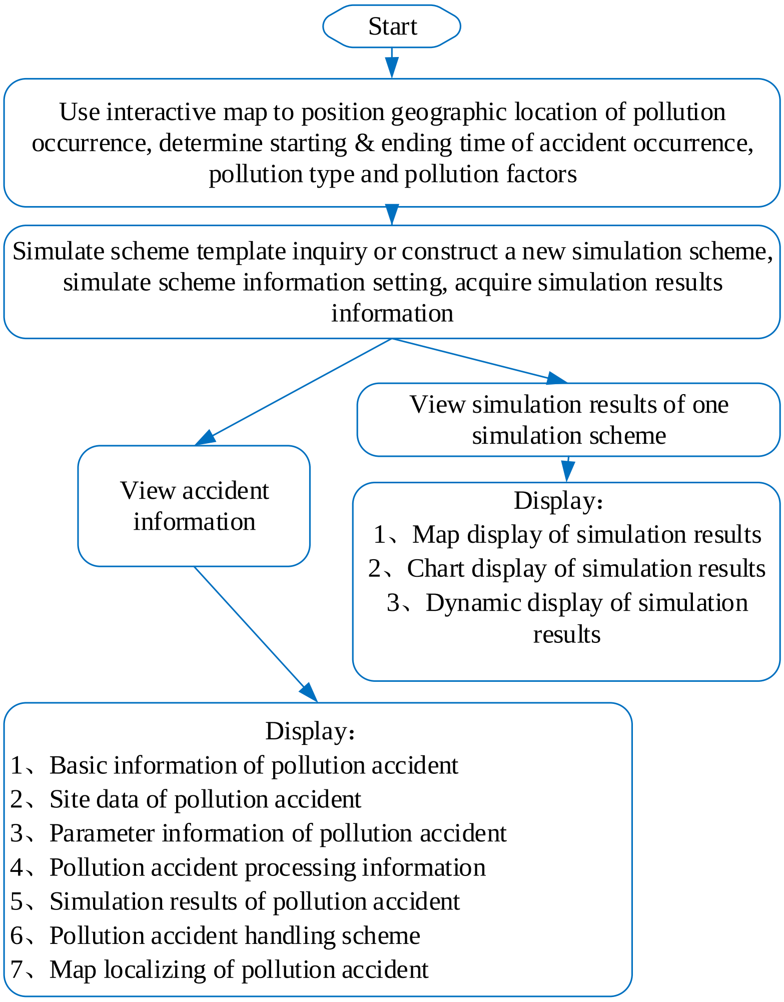


Figure 6-9. Emergency pollution simulation flowchart.

6.4 Construction of water quality SEWS prototype and implementation example of water quality simulation and early warning

6.4.1 Software & hardware environment constructed by the system

（1）Deploy server environment

1）Elastic compute node server

Provided by Aliyun elastic compute server, node server cluster is ordered as required, and the most basic configuration of its single nodes is as follows:

Example specifications: one-core Intel（R）Xeon（R）CPU E5-2680 v3 @2.50GHz；RAM: 1G；(general type N1，ecs.n1.tiny）

Example series: series model II, default to I/O optimized example; SSD 40G disk

Operating system: Windows Server 2012 standard-edition 64-bit Chinese version

Bandwidth peak value: 25M

Region: Beijing

2）Master server environment

Master server: provided by Aliyun elastic compute server, its configuration is as follows:

Example specifications: 2-core Intel（R）Xeon（R）CPU E5-2680 v3 @2.50GHz；RAM: 4G；(General type N1，ecs.n1.tiny)

Example series: Series model II, default to I/O optimized example; SSD 40G disk

Operating system: Windows Server 2012 standard-edition 64-bit Chinese version

Bandwidth peak value: 50M

Region: Beijing

（2）Development environment

Windows 7 64bit，Intel（R） Core（TM）i7-2760QM CPU@2.4GHz 2.4GHz, desktop workstation of 10G RAM; SQLServer 2012 database; ArcGIS geographic information system software; VMware Workstation 10 virtual machine; Microsoft Visual Studio .NET 10.0.

6.4.2 **Interfaces of Mudan River water quality SEWS prototype system**

Besides some above mentioned key functions, Mudan River water quality early warning system also have some other auxiliary functions such as data acquisition, thematic map, section monitoring and bulletin generation and so on. Its interfaces are shown in Figure 6-10.

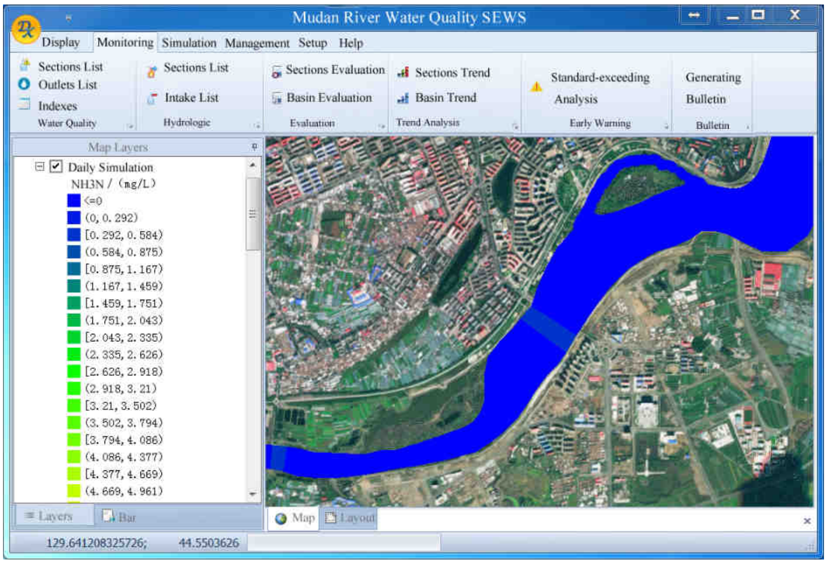


Figure 6-10. Interface of Mudan river water quality SEWS.

This system is distributed architecture which has elastic node management tool and master server management tool, its interfaces are respectively shown in Figure 6-11 and Figure 6-12. Node server tool is used to set IP address, communication port and so on which connect nodes and master server, master server management tool can monitor working status of elastic nodes and make statistics of quantity executed by the model, average inflow data volume, simulation task load which is being processed, maximum response time, etc.

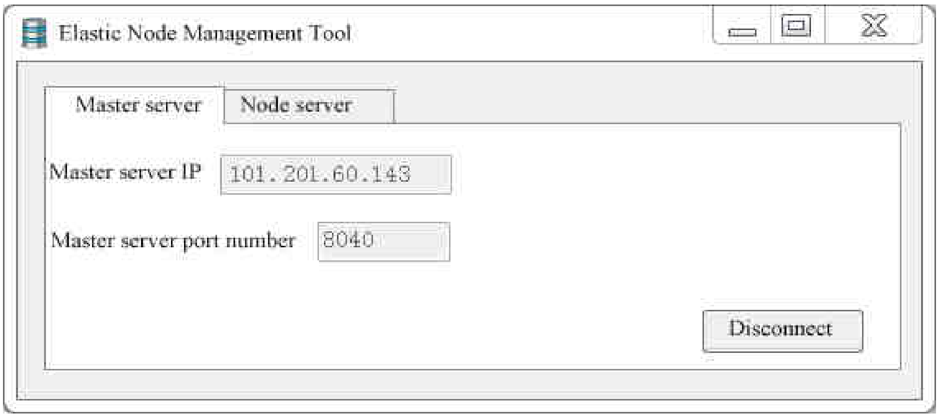


Figure 6- 11. Elastic node management tool.

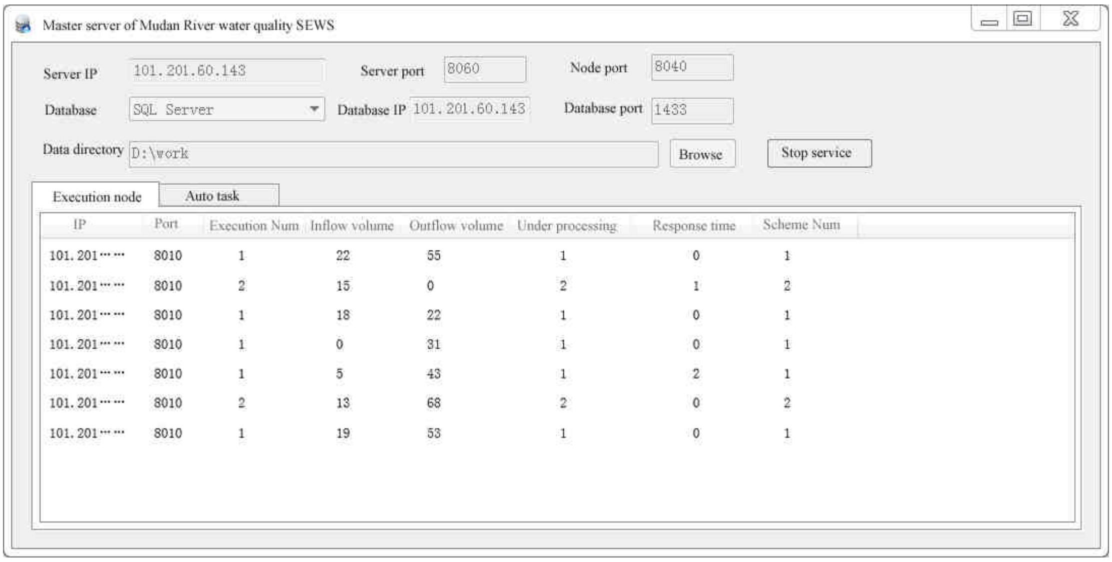


Figure 6-12. Master server management tool.

6.4.3 Scenario analysis of daily water quality simulation in open-water period

Water quality simulation and forecasting scenario analysis is an application of Mudan River water quality model constructed in Chapter 3 and it predicts spatial-temporal distribution which may occur in the future water quality according to assumptions and then realizes the goal of forecasting and early warning analysis.

（1）Scenario case setting

We now conduct a scenario simulation of daily water quality process under different inflow assurance rates of Mudan River mainstream. We assume that simulation period is from June to August of 2017 when the watercourse is in open-water period, inflow assurance rates are respectively 75% (slightly dry), 50% (normal) and 25% (slightly abundant). Flow quantities of different inflow assurance rates are shown in Table 6-1, simulation index is COD, model calculation step length is 6 s, output results step length is 24 h. Roughness efficient is 0.035, decay coefficient is 0.03/day, initial water depths are respectively 2.25 m, 2.49 m and 2.82 m, initial concentrations adopt measured COD (Table 6-2) values on water quality monitoring sections of Xige, Wenchun Bridge, Jiangbing Bridge and Hualin Bridge in May of 2014 in order to conduct interpolation of all grids, main sewage outlets along the riverbank include Ningan municipal drainage exit, Wenchun Town domestic sewage drainage exit, Hailang River import entrance, Hengfeng Paper drainage exit, Beian River import entrance, Mudanjiang urban sewage plant drainage exit, Caihe Linhai Paper drainage exit and Caihe Town domestic sewage drainage exit. Flow quantities of Xige section and Hailang River import section adopt flow quantities respectively under 75% , 50% and 25% inflow assurance rates, and flow quantities of sewage outlets adopt flow values converted from sewage discharge in 2014.

Table 6-1. Flow Quantities and Boundary Conditions of Scenario Case of COD simulation.

No.	Names of Section and Drainage Exit	Number of Days	Flow Quantities under Different Inflow Assurance Rates (m^3/s)		
			75%	50%	25%
1	Xige Section	0	56.46	99.66	162.62
		30	56.46	99.66	162.62
		31	71.2	120.52	188.90
		90	71.2	120.52	188.90
2	Ning'an Municipal	0		0.23	
		90		0.23	
3	Wenchun Town Domestic Sewage	0		0.08	
		90		0.08	
4	Hailang River Entrance	0	27.39	47.74	79.87
		30	27.39	47.74	79.87
		31	33.6	58.19	94.88
		90	33.6	58.19	79.87
5	Hengfeng Paper	0		0.1	
		90		0.1	
6	Bei'an River Import	0		0.047	

	Entrance	90	0.047
7	Urban Sewage Plant	0	1.18
	Drainage Exit	90	1.18
8	Linhai Paper	0	0.04
		90	0.04
9	Caihe Town Domestic	0	0.09
	Sewage	90	0.09

Table 6-2. Measured COD concentration values of water quality monitoring sections in May of 2014.

No.	Section Name	Concentration (mg/l)
1	Xige	4.7
2	Wenchun Bridge	4.8
3	Hailang	5.3
4	Riverside Bridge	4.6
5	Caihe Bridge	6.1

（2）System Implementation Process

We construct scenario simulation scheme by scheme tool newly constructed by the system, when the system starts operating the scheme, backstage operations of the system are successively: 1) Taking quantity of model simulation schemes currently executed by each execution node as the judgment basis, we find the optimal node with least execution number. 2) Inquire local scheme files through scheme ID. 3) Read data of local scheme and transmit to the optimal execution node. 4) After the execution node receives the scheme, it will store into corresponding catalogue and conduct unique identification by naming files with scheme ID. 5) This execution node will make one copy of Mudan River water quality model files from template and conduct unique identification by naming file folders with scheme ID. 6) This execution node reads scheme data and substitute it for copied Mudan River water quality model files, 7) This execution node tries to operate this Mudan River water quality model after replacement in multithreading. 8) If the operation is successful, then the route where this scheme is located will be registered into distributed file system. If among previous step 3) to 7), error information is returned, the system will eliminate this node and return to step 1) until the scheme is successfully operated. 10) After execution node operation is finished in water quality simulation model, this execution node will register completed logo files completed in model operation into distributed file system. 11) Download simulation results, through the distributed file system, we will inquire which route of which elastic execution node the simulation data of the scheme is located, and then we conduct corresponding downloading. 12) After simulation results files are downloaded, we'll conduct reading in GIS client side and conduct spatial-temporal dynamic display, and simulation results at different time are shown in Figure 6-13 and Figure 6-15. Example codes implemented by the whole process are shown in Appendix I.

（3）Analysis of Simulation Results

Figure 6-13 distribution diagram of COD concentration in urban area on the 1^{st}, 30^{th}, 60^{th} and 90^{th} day under 75% inflow assurance rate. From the figure we can observe spatial distribution situation of water quality concentration of inflows from mainstream and Hailang River. After the two inflows are mixed together, COD basically reaches uniform mixing on section nearby downstream through a certain-distance relocation diffusion.

Table 6-3 is a comparison of minimum COD concentration values of water quality monitoring sections under different inflow assurance rates, and it can be found from the table that minimum COD concentration values of all water quality monitoring sections somehow decreases with increasing inflow quantity, which indicates that on the condition that discharge capacity and pollutant concentration at sewage outlets stay unchanged, the greater inflow quantity from mainstream and branches is, the lower pollutant concentration of mainstream will be. From the whole simulation period, variation range of COD concentration is among 4-6 mg/l, and water quality type belongs to grade III water. Simulation results of COD concentration variation of mainstream under 50% and 25% inflow assurance rates are similar to those under 75% inflow assurance rate as shown in Figure 6-14 and Figure 6-15.

Table 6-3. Comparison of Minimum COD concentration values of water quality monitoring sections under different inflow assurance rates (mg/l).

Section Name	Inflow Assurance Rate		
	75%	50%	25%
Wenchun Bridge	4.735	4.710	4.695
Hailang	4.675	4.630	4.610
Riverside Bridge	4.670	4.620	4.600
Caihe Bridge	4.935	4.770	4.680

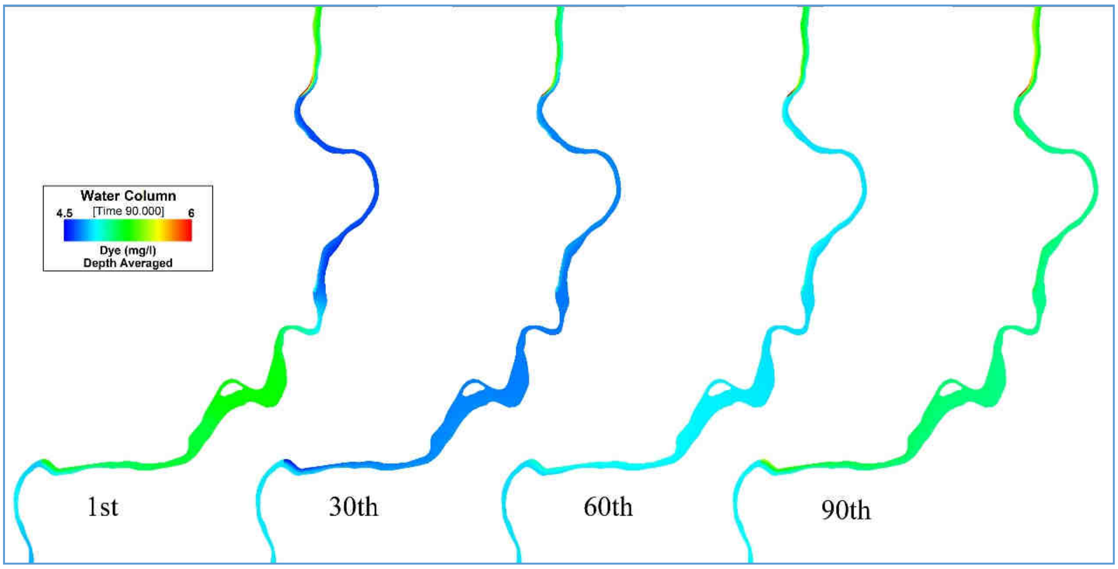


Figure 6-13. COD concentration variation at urban segment under 75% inflow assurance rate.

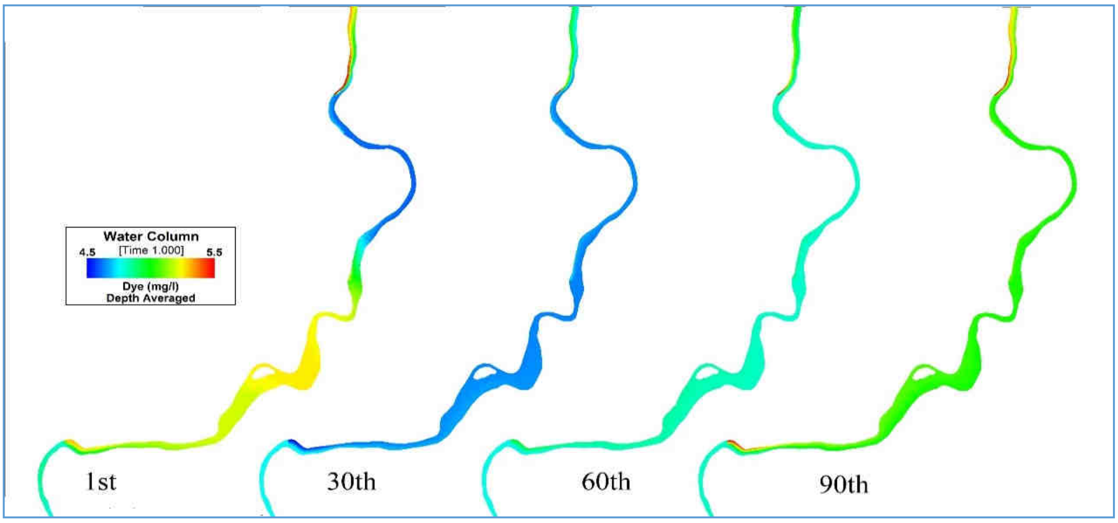


Figure 6-14. COD concentration variation at urban segment under 50% inflow assurance rate.

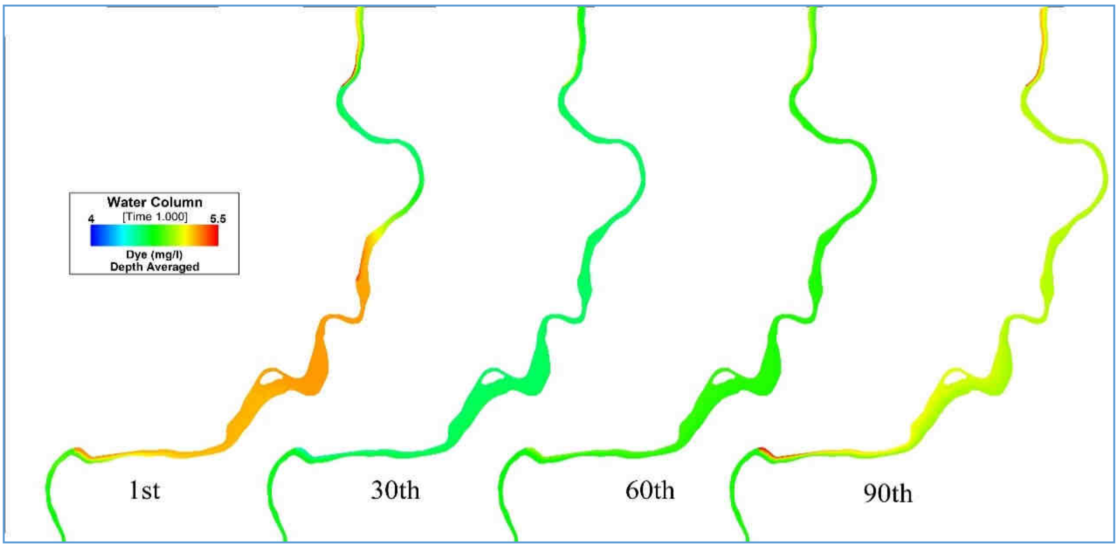


Figure 6-15. COD concentration variation at urban segment under 25% inflow assurance rate

6.4.4 Scenario analysis of emergency water pollution in ice-covered period

（1）Scenario Case Setting

We now conduct a scenario simulation of water quality process in Mudan River mainstream sudden pollution accident. We assume that chemicals vehicle overturn pollution accident occurred on Wenchun Bridge at some time, 20 tons of NH_3N pollutants are instantaneously discharged into watercourse and fully mix with inflow from watercourse rapidly, inflow assurance rates in mainstream and Hailang River are respectively set as 75% (slightly dry), 50% (normal) and 25% (slightly abundant), initial water depths are respectively set as 1.95 m, 2.06 m and 2.21 m, pollutant attenuation coefficient is 0.01/day, and output step length is 2 h.

（2）Analysis of Simulation Results

Figure 6-16 is a schematic diagram of pollutant variation status at urban segment under different inflow assurance rates, Figure 6-17 is about pollutant variation process curves of major sections under 75% inflow assurance rate, major sections include two water sources namely Railway water source and West water source; three water quality monitoring sections namely Hailang section, Jiangbing Bridge section and Caihe Bridge. It can be observed from Figure 6-17 that under 75% inflow assurance rate, when overturn accident of chemicals vehicle occurred on Wenchun bridge, it takes about 0.67

day namely 16h for pollution group to arrive at Railway water source, pollutant concentration of section after 18.5h reached the maximum value, being 4.9 mg/l; it takes about 18h for pollution group to arrive at west water source, and pollutant concentration of section after 21h reaches the maximum value, being 4.35 mg/l ; it takes about 19h for pollution group to arrive at Hailang section, and pollutant concentration of section after 22h reaches the maximum value, being 4.3 mg/l； it takes about 22h for pollution group to arrive at Jiangbing Bridge section, and pollutant concentration of section after 26h reaches the maximum value, being 3.85 mg/l; it takes about 62h for pollution group to arrive at Caihe Bridge section, and pollutant concentration of section after 76.8h reaches the maximum value, being 1.61 mg/l. Spatial distribution of pollutant concentration at urban segment under 75%, 50% and 25% inflow assurance rates is shown in Figure 6-16.

Pollutant relocation diffusion process under 50% and 25% inflow assurance rates is similar to that under 75% inflow assurance rate. Pollutant variation characteristics under different inflow assurance rates are shown in Table 6-4. It can be found from the table that as inflow quantity increases namely inflow assurance rate increases, propagation of pollution group to downstream is accelerated. It respectively takes 16 h, 15.5 h and 14.6 h for pollution group to arrive from Wenchun Bridge at Railway water source under inflow assurance rates of 75%, 50% and 25%, and 62 h, 58.8 h and 54 h for it to arrive at Caihe Bridge. In addition, concentration peak value of pollution group to arrive at one section slightly decreases as inflow assurance rate increases.

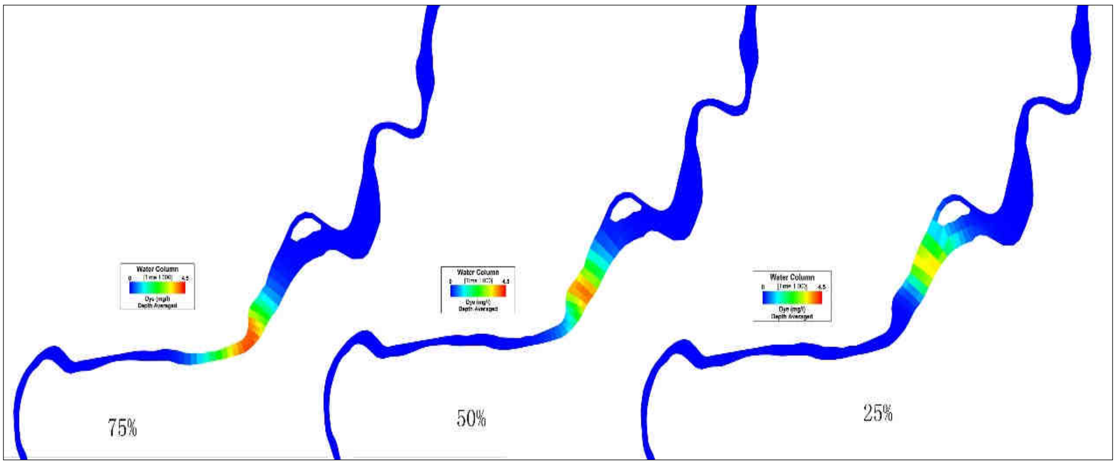


Figure 6-16. Spatial distribution state of pollutants in urban segment under different inflow assurance rates.

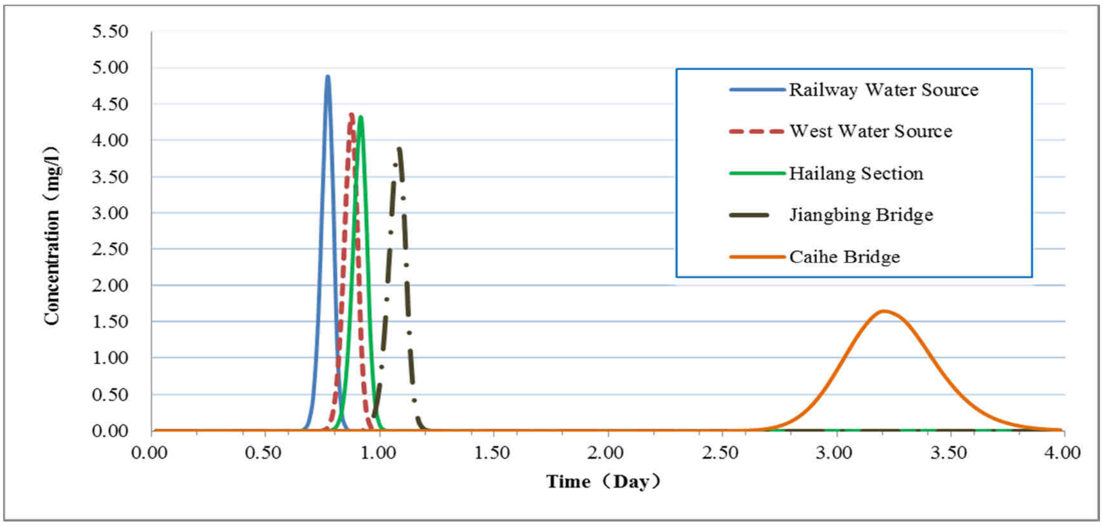


Figure 6-17. Pollutant concentration variation curves under 75% inflow assurance rate.

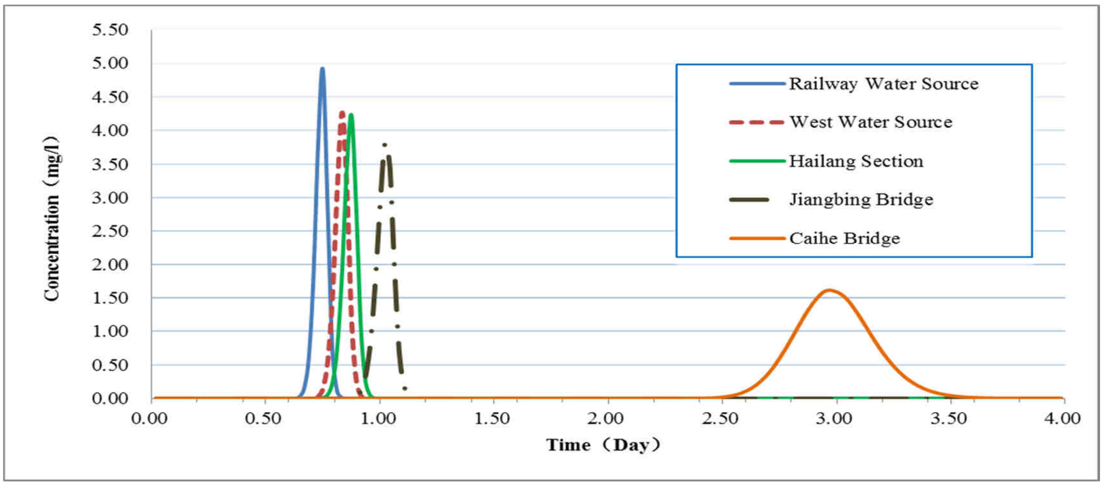


Figure 6-18. Pollutant concentration variation curves under 50% inflow assurance rate.

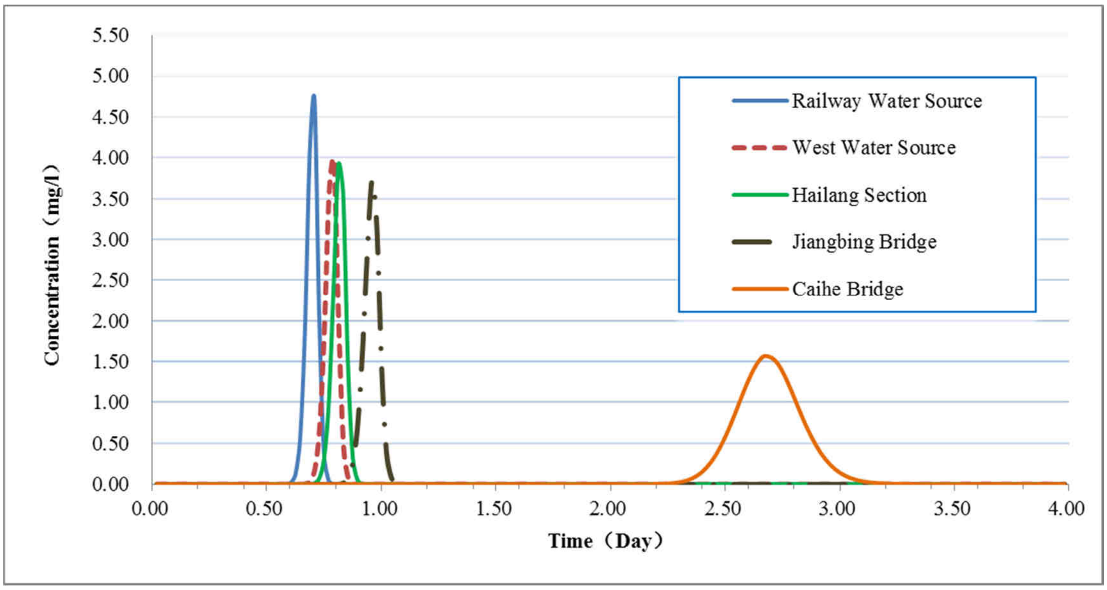


Figure 6-19. Pollutant concentration variation curves under 25% inflow assurance rate.

Table 6-4. Statistical table of pollutant variation characteristics on major sections under different inflow assurance rates.

	Section Name	Railway Water Source	West Water Source	Hailang Section	Riverside Bridge	Caihe Bridge
75% (Diachrony)	Forward Peak Value	16 18.5	18 21	19 22	22 26	62 76.8
	Concentration	4.9	4.35	4.3	3.85	1.61
50% (Diachrony)	Forward Peak Value	15.5 18	17.3 19.9	18.2 21	21 24.5	58.8 70.8
	Concentration	4.85	4.25	4.2	3.77	1.61
25% (Diachrony)	Forward Peak Value	14.6 17	16.3 19	17 19.5	19.9 23	54 63.8
	Concentration	4.8	4.1	4.1	3.75	1.56

6.4.5 Distributed computing efficiency test

The servers deployed in Ali elastic node cluster are used to conduct performance test, in order to simplify complicacy of the test, servers of all elastic compute nodes which are exactly like the most basic configuration described in Section 6.4.1. When single-node mode is used to conduct forecasting simulation, each scheme monopolizes servers, and total time consumption is the sum of time consumptions of these schemes, namely 4.48h. However, when distributed model compute mode is used, total time consumption is decided by the scenario simulation scheme with the longest time consumption, so its total time consumption is 1.1579h. Distributed compute mode greatly improves computing efficiency of scenario simulation. When this mode is adopted and program planning of scenario simulation changes, number of node servers can be increased or reduced as required, which saves usage cost of hardware equipment of servers on the condition that simulation and forecasting efficiency is improved (Table 6-5). If simulated and forecasted watercourse scope is broader, time period is longer, and there are more model grids, advantages of distributed model computation will be more obvious.

Table 6-5. Computing time consumption test of water quality scenario simulation.

Compute Mode	Scenario Simulation of Daily Water Quality Forecasting (h)			Scenario Simulation of Emergency Water Quality Forecasting (h)			Total Time (h)
	75%	50%	25%	75%	50%	25%	
Single Node	1.1579	1.4408	1.4438	0.1469	0.1455	0.1451	**4.48**
Distributed System	1.1579	1.4408	1.4438	0.1469	0.1455	0.1451	**1.1579**

6.5 Chapter conclusion

This chapter designed the logic structure and function framework of water quality warning system, logically divided water quality warning system into physical layer, data layer, application layer and presentation layer, and respectively elaborated them. After that, this chapter further discussed the realization methods of several key technologies relevant to early warning system including the database of the system, water quality evaluation, analysis on water quality tendency, early warning of exceeding standard of water quality, normal water quality simulation as well as sudden pollution forecasting and early warning. Normal water quality simulation and sudden pollution forecasting and early warning were core function of SEWS based on hydrodynamic water quality model. And the prototype of the water quality early warning system, elastic node management tool and master server management tool were also introduced in this chapter. Finally, this chapter elaborated the application scenarios by constructing both the normal and the sudden pollution forecasting and early warning scenario simulation, and revealed advantage of distributed mode efficiency, and implementation method and application efficiency of distributed early warning system of water quality simulation.

Chapter 7

Conclusions and prospects

7.1 Conclusions

Objective of this study is to establish a powerful modeling system for forecasting and early warning of water quality dynamics in Mudan River. Specific goals in the establishment of the system include a complete list of functions required for daily and accidently administration and management of water quality in the river and an efficient computation to the distributed simulation with thousands of nodes for simultaneously numerical solution. As seen in the previous chapters, this objective and its relevant goals have been successfully achieved. Several critical technical problems have been deeply examined and solved for the establishment. Efforts have been devoted to the following three aspects: (1) construction of a water quality model for Mudan River and validation of the required key parameters in both the ice-covered and open-water periods of the river; (2) implementation of the distributed computations and the related computational problems to establish the water quality model; (3) realization of such key technologies of SEWS as water quality evaluation and water quality trend analysis under GIS environment. Main conclusions on the three aspects can be summarized as follows:

(1) Construction of a water quality model for the Mudan River and validation of related parameters

A 2D hydrodynamic water quality model was established for the representative river in cold North China, Mudan River according to watercourse terrain data and water quality monitoring data. COD and NH_3N concentrations in the ice-covered and open-water periods were simulated. Moreover, diffusion coefficient, bed roughness, comprehensive rate of decay and other parameters were calibrated and verified. Based on analysis of research results, we concluded that:

1) The established water quality model was used to simulate COD and NH_3N concentrations in urban trunk stream of the Mudan River, which reported high simulation accuracy. According to research results, simulation errors of COD concentration on four validation sections ranged between 5.86%~18.43% and simulation errors of NH_3N concentration varied between 14.88% and 39.58%. Simulated results can objectively reflect transportation of COD and NH_3N in this section during the ice-covered and open-water periods, indicating that the model is feasible to make water quality simulation and early warning for the trunk stream of the Mudan River. Meanwhile, Mudan River is a typical river in cold North China. Therefore, the established model can be used in spatial-temporal dynamic simulation and forecasting of hydrodynamic water pollution of rivers around the North China.

2) Rates of decay of COD and NH_3N in the ice-covered period were lower than those in the open-water period. Results showed that the rates of decay of COD and NH_3N in the trunk stream of Mudan River during the open-water period were 0.03/day and 0.05/day, while those during the ice-covered period were 0.01/day and 0.02/day. Air temperature, upstream inflow and ice cover are main influencing factors of rate of decay.

3) Roughness is an important influencing factor of the river hydrodynamic process. Watercourse roughness during the open-water period only considers bed roughness, but it also involves ice cover resistance during the ice-covered period. When establishing the SEWS, it is necessary to configure functional interfaces of related parameters during the ice-covered and open-water periods parameters.

(2) Implementation of distributed computations and related computational problems of the established water quality model

Due to compute-intensive process in the water quality model, it has high requirements on computing resources and water quality simulation based on common serves is time-consuming. A single server couldn’t implement abundant simulation schemes in order when the SEWS has many clients, large simulation scope, fine meshing, small step length or plenty water quality indexes.

Therefore, this study put forward a distributed water quality model calculation method which distributes computations of water quality simulation onto different elastic nodes. In daily water quality simulation and sudden simulation programs, pollution concentrations and hydrologic data at monitoring stations have to be set, which could be read from databases. Since simulation of the water quality model is presented as programs, water quality and hydrologic data under the GIS environment are fixed in programs, which means that water quality simulation is composed of water quality simulation planning and water quality simulation computing. Both reflect characteristics of distributed computations.

Under this circumstance, this study discussed the general framework of the distributed SEWS, interaction between water quality model and GIS, server model design of the distributed early warning system, technical requirements and implementations of distributed architecture on distributed file system, distributed computations of water quality model, and load balancing scheduling method in the computational process. It proposed to store water quality simulation programs as binary files to reduce data transmissions under distributed environments and realize the method for share of simulation program files and storage of simulated results.

(3) Realization of related core technologies of SEWS

As a practical water quality SEWS, implementation methods of other core functions shall be considered except for establishing a water quality model, solving computation expansibility of the model and shortening running time by increasing parallel operations. As a result, this study also discussed establishment method and content of the water quality early warning database, analysis on implementation method of water quality evaluation technique, analysis on construction of water quality tendency technique, construction approach of early warning technique, and implementation approaches for conventional water quality and emergency water pollution simulation analysis.

The water quality early warning database was established by combining the spatial database and traditional relational database, which could accommodate different types of data. Water quality evaluation based on the most unfavorable factor was analyzed. The implementation algorithm of water quality tendency analysis was analyzed using the seasonal Kendall method. Considering different types of early warning, the early warning processes based on client and server were analyzed, respectively. Early warning based on the client can be controlled manually and simulates the forecasting and early warning approach repeatedly. The server make early warning of water quality based on automatic monitoring section. Both conventional and emergency water quality simulation and forecasting are based on the established model. Interfaces for related parameter setting were built up under the support of GIS. According to model analysis, GIS can read and make a visual spatial-temporal dynamic display of the analysis.

7.2 Innovations

The study mainly have the following two scientific innovations:

(1) A 2D modelling system was established in the study for water quality forecasting and early warning in Mudan River.

This modelling system has a powerful functions for simulation of water quality dynamics in Mudan River. It considers the different water pollution transmissions in the river during both ice-covered and open-water periods. Through simulations of COD and NH_3N concentrations, the following key parameters were verified for the system: convective diffusion coefficient, bed roughness and comprehensive rate of decay at different periods of the river. This verification can be applied to monitor such pollution indexes as DO, TP and TN in Mudan River.

(2) An efficient architecture was proposed in the study for computation of the system

This architecture is with a structure of distributed computation for modelling water quality dynamics in the river. Its major components include the distributed compression, storage, sharing and download of simulation programs as well as dynamic elastic distribution and computations of the water quality model. It can improve distributed computations and extendibility of the SEWS, demand supply of server resources, and emergency responses and handling capability to water pollution events.

7.3 Research prospects

Limited by time and energy, many contents in this study need to be further improved, mainly including:

(1) Establishment of the water quality simulation and early warning model can consider more complicated factors

Take NH_3N for example. Non-point source pollutions (e.g. water and soil loss, return water of farmland irrigation, large scale farming and rural garbage) and urban domestic sewage are main sources of NH_3N in the study area. However, there’s no observed data of non-point source pollutions, which explains the lower simulation accuracy of NH_3N than COD. Hence, data related with non-point source pollutions in the Mudan River shall be further explored to comprehend pollution law of NH_3N completely.

In cold areas, ice-covered period includes ice period, freeze-up period and frostless period. The comprehensive roughness and pollutant propagation law of the water quality model in different periods are different. For example, comprehensive roughness increased gradually during the ice period and became stable until the complete freeze-up period and then decreased gradually again in the frostless period to the value of the open-water period. In fact, roughness during the ice-covered period only reflected the roughness during the freeze-up period, but didn't involve the ice period and frostless period due to data missing. Given certain conditions, 2D or 3D simulation technique of pollutant propagation in rivers during ice period, freeze-up period and frostless period in North China could be further studied.

(2) SEWS study based on GIS support

GIS has many other applications in SEWS. This research realized distributed computations and related functions of water quality model under the support of GIS, including spatial database, visualization positioning, spatial query, spatial-temporal dynamics display of water quality, and so on. After 50-year development, mature GIS products in mobile, PC, server and cloud have been launched. Supported by these products, it can construct an Internet of Things of water quality early warning system based on water quality monitoring sensors, a mobile water quality monitoring and early warning system based on portable terminal, and the network information sharing water quality early warning system based on WebGIS. I believe all of these are important research areas and have important values. These systems have been researched and developed in some river basins. River water environmental management can be further enhanced with an intelligent water environment and quality early warning, decision-making and command system which is established by combining distributed computation technique of water quality model, private cloud, public cloud and intelligent computation.

(3) Improve water quality simulation and its visualization effect based on GPU parallel computing technique

Water quality simulation research based on physics has two technical routes. One is based on traditional CPU algorithm and the other one is based on GPU (Graphics Processing Unit) technology. Conceptualization of related water environment is reduced as much as possible based on the virtual reality technology of GPU. Water environment is adjusted and optimized on the GPU platform by available hydrodynamic model and water quality model, which is convenient for making simulation experiment intuitively, visually and accurately.

Different from the traditional research idea, simulation based on GPU respects the mechanism of natural law more and is easier to be used by end users. It can provide a virtual platform for intuitive water quality simulation, thus enabling to further improve the model and optimize water quality.

Although there are plenty application researches of water quality model and theoretical researches have achieved great development, it will encounter many problems in practical applications. These problems are caused by various reasons. 1) Modeling requires mass data. Integrity, systematicness and authenticity of data all can influence simulation accuracy of the model directly. The model involves a lot of complicated parameters. Modeling, parameter calibration and validation are all effort-consuming. Uncertainty of parameters will cause inaccurate results. 2) Mechanism of water pollution simulation still remains unknown and many water pollution simulations couldn't be expressed by mathematical formulas. Certain conceptualization is needed during modeling, resulting in model distortion. 3) Due to influences from artificial subjective factors, the model has big simulation errors and simulated results are not intuitive and clear enough. These can be solved by GPU programming algorithm, but it has high requirements on accuracy of parameters related with river bed and pollution propagation. Given adequate data, it can further explore implementation methods of the distributed SEWS based on GPU parallel computation and 3D simulation technology, and enhance large-scale fine water quality simulation, early warning and computing capability as well as visualization of simulations.

Reference

[1] Afrati F N, Ullman J D. Optimizing joins in a map-reduce environment Proceedings of the 13th International Conference on Extending Database Technology, 2010, pp. 99-110.

[2] Ai H N, Zhang W S, Hu X B. The research and application progress of environmental fluid dynamics code. Journal of Water Resources Research, 2014, 03, pp. 247-256. (In Chinese)

[3] Ambrose R B, Wool T A, Connolly J P. WASP4, A hydrodynamic and water quality model-model theory, User's Manual and Programmer's Guide, US Environmental Protection Agency, Athens, Ga, USA, 1988.

[4] Ambrose R B, Wool T A, Martin J L. The water quality analysis simulation program WASP5 model documentation and user manuals. Environment Research Lab USEPA, Athens, Geogia, USA, 1993.

[5] Arakawa A, Lamb V R. Computational design of the basic dynamical processes of the UCLA general circulation model. Methods Comput. Phys. 1977, 17, pp. 174–265.

[6] Aron R H, Aron I M. Statistical forecasting models (I): control monoxide concentrations in the Los Angeles basin. Journal of Air Pollution Control Association, 1978, 28, pp. 681-692. (In Chinese)

[7] Artioli Y, Bendoricchio G, Palmeri L. Defining and modelling the coastal zone affected by the Po river (Italy), Ecological Modelling, vol. 184, no. 1, 2005, pp. 55–68.

[8] Bai J H, Xiao R, Cui B C, et al. Assessment of heavy metal pollution in wetland soils from the young and old reclaimed regions in the Pearl River Estuary, South China, Environmental Pollution, vol. 159, no. 3, 2011, pp. 817–824.

[9] Bencala K E. Walters R A. Simulation of solute transport in a mountain pool-and-riffle stream: a transient storage model, Water Resources Research, vol. 19, no. 3, 1983, pp. 718–724.

[10] Berthold J, Dieterle M, Loogen R. Implementing parallel google map-reduce in eden. Proceedings of the 15th International Euro-Par Conference on Parallel Processing. Springer-Verlag, 2009, Vol.5704, pp.990-1002.

[11] Bode H, Nusch E A. Advanced river quality monitoring in the Ruhr basin. Water Science & Technology. 1999, 40(10), pp. 145-152.

[12] Bostock H H, Richards P A, Rogers M C, Waldon M G. A GIS approach to potable water protection on the lower mississippi river. AWRA: Resource Development of the Lower Mississippi River.1991.

[13] Boussinesq J. Theorie analyque de la chaleur; Gathier-Villars: Paris, France, 1903.

[14] Brown L C, Barnwell T O. The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: Documentation and user manual; EPA/600/3-87/007; Environmental Research Laboratory, USEPA: Athens, GA, USA, 1987.

[15] Burn D H, Mcbean E A. Optimization modeling of water quality in an uncertain environment, Water Resources Research,vol. 21, no. 7, 1985, pp. 934–940.

[16] Camp T R. Water and its impurities, Reinhold, New York, NY, USA, 1963.

[17] Cerco C F, Cole T. User's guide to the CE-QUAL-ICM three dimensional eutrophication model, Release Version 1.0, Technical Report EL-95-15,US Army Eng. Waterways Experiment Station, Vicksburg, MS, USA,1995.

[18] Chambers F B, Duce D A, Jones G P. Distributed computing. Encyclopedia of Biometrics, 1984,

28(2), pp. 137-146.
[19] Chambers P A, Brown S, Culp J M, et al. Dissolved oxygen decline in ice-covered rivers of northern Alberta and its effects on aquatic biota. Journal of Aquatic Ecosystem Stress and Recovery, 2000, 8(1), pp. 27-38.
[20] Chapra S C, Pelletier G J, Tao H. QUAL2K: A modeling framework for simulating river and stream water quality, Version 2.11: Documentation and user's manual; Department of Civil and Environmental Engineering, Tufts University: Medford, OR, USA, 2008.
[21] Chapra S C, Pelletier G J. Qual2k: a modeling framework for simulating river and stream water quality. Kenyon Review, 2012, 34(2), pp. 119-120.
[22] Chen B Q, Tan D B, Song L. Application research of gis technology in suddenly-happened water pollution accident emergency response system. Journal of Yangtze River Scientific Research Institute, 2010, 27(1), pp. 29-32. (In Chinese)
[23] Chen F. Analysis of the effects of process variations on delta morphology and stratigraphy in Delft3D computational models. Delft University of Technology, the Netherlands, 2014.
[24] Chen J, Lu H Y, Yao D D. An algorithm for Map/Reduce dynamic loading balancing policy based on processing time. Journal of Jiangxi Normal University: Natural Sciences Edition, 2013, 37(1), pp. 51-55. (In Chinese)
[25] Chen S S, Fang L G, Li H L, Zhang L X. Saltwater intrusion analysis experiential model for Pearl River Estuary, South China: A case study in Modaomen watercourse. Adv. Water Sci. 2007, 18, pp. 751–755. (In Chinese)
[26] Chui X D, Liu B C. Research and control of Mudanjiang River non-spot source water environment pollution. North. Environ. 2005, 3, pp. 70–71. (In Chinese)
[27] Cole T M, Buchak E M. CE-QUAL-W2: a Two-dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model, Version 2.0. Department of Civil and Environmental Engineering, Portland State University, Portland, 1995.
[28] Danish Hydraulic Institute, MIKE 3 eutrophication module, User Guide and Reference Manual, Release 2. 7, Danish Hydraulic Institute, Horsholm, Denmark, 1996.
[29] Danish Hydraulic Institute, MIKE21: User guide and reference manual, Danish Hydraulic Institute, Horsholm, Denmark, 1996.
[30] Danish Hydraulics Institute, MIKE11, User guide & reference manual, Danish Hydraulics Institute, Horsholm, Denmark, 1993.
[31] Debolskaya E I, Isaenkov A Y. Mathematical modeling of the transportation competency of an ice covered flow. Hydrophys. Process. 2010, 37, pp. 558–567.
[32] Delft3D-FLOW, Simulation of multi-dimensional hydrodynamic flows and transport phenomena, including sediments – user manual. The Netherlands, 2006.
[33] Deltares. Delft3D-FLOW user manual: Simulation of multi-dimensional hydrodynamic flows, 2013.
[34] Diehl P, Gerke T, Jeuken A. Early warning strategies and practices along the river Rhine. Handbook of Environmental Chemistry, 2006, 5, pp. 99-124.
[35] Ding X R, Xu J, Yao Q. GIS and numerical model integrated for space-time simulation of sudden water pollution. Journal of Hohai University (Natural Sciences), 2003, 31(2): 203-206. (In Chinese)
[36] Dobbins W E. BOD and oxygen relationships in streams, Sanitary Engineering Division, American Society of Civil Engineers, vol. 90, no. 3, 1964, pp. 53–78.

[37] Druon J N, Mannino A, Signorini S, Mcclain C, Friedrichs M, Wilkin J, Fennel K. Modeling the dynamics and export of dissolved organic matter in the northeastern U.S. continental shelf. Estuar. Coast. Shelf Sci. 2010, 88, pp.488–507.

[38] Esterby S R. Review of methods for the detection and estimation of trends with emphasis on water quality applications, Hydrological Processes, vol. 10, no. 2, 1996, 127–149.

[39] European Commission. Implementing the geographical information system elements (GIS) of the Water Framework Directive (WFD Guidance Document No. 9), 2015.

[40] Fan C, Ko C H, Wang W S. An innovative modeling approach using Qual2K and HEC-RAS integration to assess the impact of tidal effect on river water quality simulation, Journal of Environmental Management, vol. 90, no. 5, 2009, pp. 1824–1832.

[41] Fan S F, Feng M Q, Liu Z. Simulation of water temperature distribution in Fenhe reservoir, Water Science and Engineering, vol. 2, no. 2, 2009, pp. 32–42.

[42] Fang X, Zhang J, Chen Y, Xu X. QUAL2K model used in the water quality assessment of Qiantang River, China, Water Environment Research, vol. 80, no. 11, 2008, pp. 2125–2133.

[43] Fischer H B, List E J, Koh R C Y, Imberger J, Brooks N H. Chapter 5—mixing in rivers. In Mixing in Inland & Coastal Waters; ACADEMIC PRESS, INC.: San Diego, CA, USA. 1979, 13, pp. 104–147.

[44] Fu C, Popescu I, Wang C, Mynett A E, Zhang F. Challenges in modelling river flow and ice regime on the Ningxia–Inner Mongolia reach of the Yellow River, China. Hydrol. Earth Syst. Sci. 2014, 18, pp.1225–1237.

[45] Golomb D, Ryan D, Underhill J, Wade T, Zemba S, Atmospheric deposition of toxics onto Massachusetts Bay - II. Polycyclic aromatic hydrocarbons, Atmospheric Environment, 1997, 31, (9), pp. 1361–1368.

[46] Gough D I. Incremental stress under a two-dimensional artificial lake, Canadian Journal of Earth Sciences, 1969. 6(5), pp.1067–1075.

[47] Grayman W M, Males R M. Risk-based modeling of early warning systems for pollution accidents. Water Science & Technology A Journal of the International Association on Water Pollution Research, 2002, 46(3), pp. 41-49.

[48] Grenney W J, Teuscher M C, Dixon L S. Characteristics of the solution algorithms for the QUAL II river model, Journal of the Water Pollution Control Federation, vol. 50, no. 1, pp. 151–157, 1978.

[49] Grizzetti B, Bouraoui F, Granlund K, Rekolainen S, Bidoglio G. Modelling diffuse emission and retention of nutrients in the Vantaanjoki watershed (Finland) using the SWAT model, Ecological Modelling, 2003, 169 (1), pp. 25–38.

[50] Gullick R W, Gaffney L J, Crockett C S, et al. Developing regional early warning systems FOR US SOURCE WATERS. Journal American Water Works Association, 2004, 96(6), pp. 68-82.

[51] Guo R, Li Y B, Fu G. Controlling factors of degradation coefficient on organic pollutant in river . Journal of Meteorology and Environment, 2008, 24(1), pp. 56~59. (In Chinese).

[52] Hamrick J M. A three-dimensional environmental fluid dynamics computer code: Theoretical and computational aspects. The College of William and Mary, Virginia Institute of Marine Science, Special Report 317, 1992.

[53] Hamrick J M. Analysis of currents in the vicinity of the proposed New Carrier Pier at Newport News Ship building: A report to Lockwood Greene engineers. The College of William and Mary, Virginia Institute of Marine Science, Gloucester Point, VA, 1992.

[54] Hamrick J M. Analysis of mixing and dilution of process water discharged into the Pamunkey river: A report to the Chesapeake Corp. The College of William and Mary, Virginia Institute of Marine Science, Gloucester Point, VA, 1991.
[55] Hamrick J M. Application of the EFDC, Environmental fluid dynamics computer code to SFWMD water conservation area 2A: A report to South Florida water management district. JMH-SFWMD- 94-01. Williamsburg, VA, 1994.
[56] Hamrick J M. Preliminary analysis of mixing and dilution of discharges into the Yorkriver: A report to the Amoco Oil Co. The College of William and Mary, Virginia Institute of Marine Science, Gloucester Point, VA, 1993.
[57] Hamrick J M. User's manual for the environmental fluid dynamic computer code. The College of William and Mary, Virginia Institute of Marine Science, Special report328, 1996.
[58] Hao Y, Sun W G, Yu Z B. Study on pollution of various sections in mudanjiang river basin by equal standard pollution load method. Modern Agricultural Science and Technology, 2013, 16, pp.217-218,220.
[59] Harmick J M. A Three-dimensional environmental fluid dynamics computer code: theoretical and computational aspects. The College of William and Mary, Virginia Institute of Marine Science: Williamsburg, VA, USA, 1992.
[60] Hou S. Establishment of the forecasting system of trans-boundary heavy pollution accidents based on GIS. Department of Env. Sci. &Eng. Harbin Institute of Technology, 2010.
[61] Howland W E, Jr T H. Pollution load capacity of streams. Water & Sewage Works, 1949, 96(7), pp. 264-266.
[62] Hu G H, Tang Z W, Xiao X Q. Trend analysis of water quality of Sanmenxia reservoir of the yellow river. Geography and Geo-information Science, 2004, 20(3), pp.86-88. (In Chinese)
[63] Huang R, Zhang H, Peng H. Research of water quality early-warning system in suzi river based on GIS and mathematic modeling. Yellow River, 2013(11), pp.70-72. (In Chinese)
[64] Huang W, Liu X, Chen X. Numerical modeling of hydrodynamics and salinity transport in Little Manatee River. Journal of Coastal Research, 2008, 10052(1):13-24.
[65] Huang Y K, Li Y P, Qiu L. Risk prediction on wharf oil spill in the lower reaches of Yangtze River based on EFDC. Water Resources Protection, 2015(1), pp.91-98. (In Chinese)
[66] Hunsaker C T, Levine D A. Hierarchical approaches to the study of water quality in rivers—spatial scale and terrestrial processes are important in developing models to translate research results to management practices, BioScience, 1995, 45(3), pp. 193–203.
[67] HydroQual Inc, 1 Lethbridge Plaza, Mahwah, NJ. USA, 2002.
[68] HydroQual Inc, A Primer for ECOMSED, Version 1.3, User's Manual, 2002.
[69] HydroQual Inc. A Primer for ECOMSED, Version 1.3; HydroQual Inc.: Mahwah, NJ, USA, 2002.
[70] Inoul T. Prediction of nitrogen oxide concentration by a regression model. Atmospheric Environment, 1986, 20, pp. 2325-2337. (In Chinese)
[71] Jeffrey D, SANJAY G. MapReduce: simplified data processing on large clusters. Communications of the ACM, 2008, 51(1):107-1
[72] Jia P, Wang Q G, Zhang D. Research on key technology in development of EFDC model based on water environmental management information system. Environmental Engineering, 2015, 33(7):131-134. (In Chinese)
[73] Jiang H Z, Shen Y M, Wang S D. Numerical simulation study on three-dimensional tidal flow and salinity in the Oujiang Estuary. Journal of Hydrodynamics, 2009, 24(1), pp. 63- 70. (In

Chinese)
[74] Kamel A. Application of a hydrodynamic mike 11 model for the euphrates river in iraq. Slovak Journal of Civil Engineering, 2008, 2, pp.1-7.
[75] Kannel P R, Kanel S R, Lee S, Lee Y S, Gan T Y. A review of public domain water quality models for simulating dissolved oxygen in rivers and streams, Environmental Modeling and Assessment, 2011, 16(2), pp. 183–204.
[76] Lesser G, Roelvink J, van Kester J, Stelling G.. Development and validation of a three-dimensional morphological model. Costal Engineering, 2004, 51, pp.883-915.
[77] Li J, Cao F F, Du G C. Study on water quality pre-warning and forecaseng system based on GIS in the Qian tang River. Zhejiang Hydrotechnics, 2008. (In Chinese)
[78] Li R Z. Advance and trend analysis of theoretical methodology for water quality forecast. Journal of Hefei University of Technology: Natural Science, 2006, 29(1), pp.26-30. (In Chinese)
[79] Li X. Modelling of eutrophic contaminants distribution and degradation processes in lake wuliangsuhai. Inner Mongolia Agricultural University, 2009. (In Chinese)
[80] Li Y P, Wang Y, Tang C Y, Anim D O, Ni L X, Yu Z B, Acharya K. Measurements of erosion rate of undisturbed sediment under different hydrodynamic conditions in Lake Taihu, China. Pol. J. Environ. Stud. 2014, 23, pp.1235-1244.
[81] Long T G, Guo J S, Feng Y Z, Huo G Y. Modulus of transverse diffuse simulation based on artificial weural network. Chong Qing Environmental Science, 2002, 24, pp.25-28. (In Chinese).
[82] Lorenz E N. Deterministic non-periodic flow, J. Atmos. Sci. 20, pp.130-141 (1963).
[83] Madala R V, Piacsek S A. A semi-implicit numerical model for Baroclinic Oceans. Journal of Computational Physics, 1977, 23, pp.167–178.
[84] Manolidis M, Katopodes N. Bed scouring during the release of an ice jam. J. Mar. Sci. Eng. 2014, 2, pp.370–385.
[85] Martin N, Mceachern P, Yu T, Zhu D Z. Model development for prediction and mitigation of dissolved oxygen sags in the Athabasca River, Canada. Science of the Total Environment, 2013, 443(3), pp.403-412.
[86] Mason R P, Lawson N M, Sullivan K A. Atmospheric deposition to the Chesapeake Bay watershed—Regional and local sources, Atmospheric Environment, 1997, 31(21), pp.3531–3540.
[87] Mellor G L, Yamada T. Development of a turbulence closure model for geophysical fluid problems. Reviews of Geophysics and Space Physics, 1982, 20, pp.851-875.
[88] Ministry of Environmental Protection of China (MEPC). 2014 China environmental state bulletin. 2015.
[89] Morley N J. Anthropogenic effects of reservoir construction on the parasite fauna of aquatic wildlife EcoHealth, 2007, 4(4), pp. 374–383.
[90] Morselli L, Olivieri P ,Brusori B, Passarini F, Soluble and insoluble fractions of heavy metals in wet and dry atmospheric depositions in Bologna, Italy, Environmental Pollution, 2003, 124(3), pp. 457–469.
[91] Mujumdar P P, Vemula V R S, Fuzzy waste load allocation model: simulation-optimization approach, Journal of Computing in Civil Engineering, 2004, 18(2), pp. 120–131.
[92] Nares C, Subuntith N, Sukanda C. Empowering water quality management in Lamtakhong River Basin, Thailand using WASP model. Res. J. Appl. Sci.: Eng. Technol. 2013, 6, pp.4485–

4491.
[93] Neto I E L, Zhu D Z, Rajaratnam N, et al. Dissolved oxygen downstream of an effluent outfall in an ice-covered river: Natural and artificial aeration. Journal of Environmental Engineering, 2007, 133(11), pp.1051-1060.
[94] NyergesT L.Coupling GIS and spatial analytic models: Spatial Data Handling .Proceedings of the 5th international symposium on spatial data handling. Charleston. South Carolina, USA.1992, 2, pp. 534-543.
[95] O'Connor D J. The temporal and spatial distribution of dissolved oxygen in streams, Water Resource Research, 1967, 3 (1), pp. 65–79.
[96] Park K, Kuo A Y, Shen J, Hamrick J M. A Three-Dimensional Hydrodynamic-Eutrophication Model (Hem-3d): Description of Water Quality and Sediment Process Submodels. Special Report in Applied Marine Science and Ocean Engineering No. 327. School of Marine Science Virginia Institute of Marine Science College of William and Mary, January 1995.
[97] Peng K, Dong X G, Zhang X F, et al. The seasonal Kendall inspection in water quality variation trend analysis at Changsha section of Xiang River. Environmental Monitoring in China, 2014(1). DOI: 10.3969/j.issn.1002-6002.2014.01.018. (In Chinese)
[98] Peng S, George Yu-zhu F U, Zhao X H. Integration of USEPA WASP model in a GIS platform. Spectrochimica Acta Part B Atomic Spectroscopy, 2010, 11(12), pp.2319–2325.
[99] Poor N, Pribble R, Greening H. Direct wet and dry deposition of ammonia, nitric acid, ammonium and nitrate to the Tampa Bay Estuary, FL, USA, Atmospheric Environment, 2001, 35, 23, pp. 3947–3955.
[100] Prowse T D. River-Ice Ecology. I: Hydrologic, Geomorphic, and Water-Quality Aspects. Journal of Cold Regions Engineering, 2001, 15(1), pp.1-16.
[101] Pu X C, Li K F, Li J, Zhao W Q. The effect of turbulence in water body on organic compound biodegradation. China Environ. Sci. 1999, 19, pp.485–489. (In Chinese)
[102] Qu M Y, Li T. Analysis of river sediment supply in middle and downstream section of Mudanjiang River mainstream. Heilongjiang Sci. Technol. Water Conserv. 2013, 41, pp.38–40. (In Chinese)
[103] Riffat R. Fundamentals of wastewater treatment and engineering, CRC Press, Boca Raton, Fla, USA, 2012.
[104] Rinaldi S, Soncini-Sessa R. Sensitivity analysis of generalized Streeter-Phelps models, Advances in Water Resources, 1978, 1(3), pp. 141–146.
[105] Roelvink J. (2006). Coastal morphodynamic evolution techniques. Costal Engineering, 53(2-3), pp. 277-287.
[106] SCC (The State Council of China). The action plan for water pollution prevention and control, 2015.
[107] Sheng Y P, Butler H L. Modeling coastal currents and sediment transport. Proceedings of 18th International Conference on Coastal Engineering, ASCE, New York, 1982, pp. 1127–1148.
[108] Sheng Y P. Finite-difference models for hydrodynamics of lakes and shallow seas. Physics-based modeling of lakes, reservoirs, and impoundments, W. G. Gray, ed., ASCE, New York, 1986, pp.146–228.
[109] Sincock A M, Lees M J. Extension of the QUASAR river water quality model to unsteady flow conditions, Journal of the Chartered Institution of Water and Environmental Management,

vol. 16, no. 1, 2002, pp. 12–17.

[110] Smolarkiewicz P K, Clark T L. The multidimensional positive definite advection transport algorithm: Further development and applications. J. Comput. Phys. 1986, 67, pp. 396–438.

[111] Smolarkiewicz P K, Grabowski W W. The multidimensional positive definite advection transport algorithm: Nonoscillatory option. Journal of Computational Physics, 1990, 86, pp.355–375.

[112] Smolarkiewicz P K, Margolin L G. On forward-in-time differencing for fluids: Extension to a curvilinear framework. Mon. Weather Rev. 1993, 121, pp. 1847–1859.

[113] Song G H, Zhang Y H. Research progress and development trend of water quality models. Equipment Environmental Engineering, 2008, 5(2), pp.32-36. (In Chinese)

[114] Song X X. Research on the key techniques of dynamic data driven early warning system for sudden river pollution accidents. Zhejiang University, 2013. (In Chinese)

[115] Sperling M V, Paoli A C. First-order COD decay coefficients associated with different hydraulic models applied to planted and unplanted horizontal subsurface-flow constructed wetlands. Ecological Engineering, 2013, 57, pp. 205-209.

[116] State Environmental Protection Administration of the P.R. China (SEPA). Environmental quality standards for surface water (GB3838-2002); China Environmental Science Press: Beijing, China, 2002 (in Chinese).

[117] Stratton F E. Nitrogen losses from alkaline water impoundments. J. Sani. Eng. Div. 1969, 95, pp.223–231.

[118] Streeter H W, Phelps E B. A study of the pollution and natural purification of the ohio river, United States Public Health Service, U.S. Department of Health, Education and Welfare, 1925.

[119] Sugihara K, Nakatsugawa M. Water quality characteristics of ice-covered, stagnant, eutrophic water bodies as influenced by climate change. Proceedings of the Japanese Conference on Hydraulics. Japan Society of Civil Engineers, 2013, pp.44-59.

[120] Sun S C,Wei H B, Xiao W H et al. Application of the freezing period hydrodynamics and water quality model to water pollution accident in Songhua River. Journal of Jilin University (Earth Science Edition), 2012, 41(5), pp. 1548-1553. (In Chinese)

[121] The U.S. Environmental Protection Agency (USEPA), Review of potential modeling tools and approaches to support the BEACH Program, Rep. No. EPA 823-R-99-002. The U.S. Environmental Protection Agency, Washington, DC, USA, 1999.

[122] Thekkath C A, Mann T, Lee E K. Frangipani: a scalable distributed file system. Acm Sigops Operating Systems Review, 1997, 31(5), pp. 224-237.

[123] Timand U S, Jolly R. Evaluating agricultural nonpoint-source pollution using integrated geographic information systems and hydrologic/water quality model, Journal of Environmental Quality, 1994, 23(1), pp. 25–35.

[124] UNDP/GEF. Development of operational tools for monitoring, laboratory and information management. UNDP/GEF Danube regional project .http:// www.undp-drp.org,2005, pp.8-28.

[125] Valett H M, Morrice J A, Dahm C N, Campana M E, Parent lithology, surface-groundwater exchange, and nitrate retention in headwater streams, Limnology and Oceanography, 1996, 41(2), pp. 333–345.

[126] Wang G, Han L X, Chang W T. Modeling water temperature distribution in reservoirs with 2D laterally averaged flow-temperature coupled model, Water Resources Protection, 2009,

25(2), pp. 59–63.
[127] Wang J P, Su B L, Jia H F,et al. Integrated model of nutrients for the Miyun reservoir and its watershed. Chinese Journal of Environmental Science, 2006, 27(7), pp.1286-1291. (In Chinese)
[128] Wang J Q, Zhong Z, Wu J. Steam water quality models and its development trend, Journal of Anhui Normal University (Natural Science), 2004, 27(3), pp. 243–247.
[129] Wang M, Li W J, Ye Z. Optimization on water environment capacity and distribution in Mudan River West Pavilion to Chai River Bridge Section. Environ. Sci. Manag. 2013, 38, pp.40–44. (In Chinese)
[130] Wang Q G, Zhao X H, Yang M S, Zhao Y, Liu K, Ma Q. Water quality model establishment for middle and lower reaches of Hanshui river, China, Chinese Geographical Sciences, vol. 21, no. 6, 2011, pp. 647–655.
[131] Wang Q, Li S, Jia P, Qi C, Ding F. A review of surface water quality models. Scientific World Journal, 2013(3), pp. 85-88
[132] Wang X E, Dong D M, Zhao W J, Li J, Zhang H L, Du Y G. Reduction mode of organic pollutants in rivers during the icebound season. Journal of Jilin University (Science Edition). 2003, 3, pp.392-395. (In Chinese)
[133] Wang Z B, Ma Y, Sun W G. Discussion about the control technologies of non-point pollution of Mudanjiang River Basin. Environ. Sci. Manag. 2011, 7, pp.3659–3662. (In Chinese)
[134] Weiler R R. Rate of loss of ammonia from water to the atmosphere. J. Fish. Res. Board Can. 2011, 36,685–689.
[135] Welander P. Wind-driven circulation in one-and two-layer oceans of variable depth, Tellus, vol. 29, 1968, pp. 1–16.
[136] Whitehead P G, Williams R J, Lewis D R. Quality simulation along river systems (QUASAR): model theory and development, Science of the Total Environment, vol. 194-195, 1997, pp. 447–456.
[137] Whitehead P G, Wilson E J, Butterfield D. A semidistributed Integrated Nitrogen model for multiple source assessment inCatchments (INCA): partI—model structure and process equations, Science of the Total Environment, 1998, 210, pp. 547–558.
[138] Wolanski E, Mazda Y, Ridd P. Mangrove hydrodynamics, Coastal and Estuarine Studies, 1992, 41, pp. 43–62.
[139] Wool T A, Davie S R, Rodriguez H N. Development of three-dimensional hydrodynamic and water quality models to support total maximum daily load decision process for the Neuse River Estuary, North Carolina. J. Water Res. Plan. Manag. 2003, 129,pp. 295–306.
[140] Wright R M, Medonell A J. In-stream de-oxygenation rate prediction. Proc. ASCE J. Env. 1979, 105, pp.323–335.
[141] Wu B G, Ding Z G. Research of Distributed Search Engine Based on Map/Reduce. New Technology of Library and Information Service, 2007,2(8), pp.52-55. (In Chinese)
[142] Wu D J, Huang Q Y, Sun H Y, et al. A water pollution diffusion model on emergency response and its visualization in a GIS platform. Geomatics and Information Science of Wuhan University, 2009, 34(2), pp.131-134. (In Chinese)
[143] Wu J. Research on classsical methods and innovatory methods for simulation and predicting river water quality illustrated with Huaihe River in Anhui province. Hefei University of Technology, 2005.

[144] Wu T S, Hamrick J M, McCutechon S C. Benchmarking the EFDC /HEM3D surface water hydrodymamic and eutrophication models: Next generation environmental models and computational methods. Society of Industrial and Applied Mathematics, Philadelphia, 1997.

[145] WU Y, HE D. Design of bottom module of server based on IOCP. Information Technology, 2007, 3, 35.

[146] Xu Z X, Lu S Q. Research on hydrodynamic and water quality model for tidal river networks, Journal of Hydrodynamics, 2003, 15(2), pp. 64–70.

[147] Yang H C, Ali D, Hsiao R L. Map-reduce-merge: simplified relational data processing on large clusters. Proceedings of the Proceedings of the 2007 ACM SIGMOD international conference on Management of data, 2007, pp.1029-1040.

[148] Yang M D, Carolyn J M. Adaptive short-term water quality forecasts using remote sensing and GIS. Advancing Water Resources Research and Management, 2007, 9(1), pp.21-33.

[149] Yih S M, DavidsonB. Identification in nonlinear, distributed parameter water quality models, Water Resources Research, 1975, 11(5), pp. 693–704.

[150] Zhang H T, Zhou C H, Wan Q. Integration of geographical information system technology and hydrological model. Advances in Water Science, 2001, 12. (In Chinese)

[151] Zhang R B, Qian X, Yuan X C, Ye R, Xia B S, Wang Y L. Simulation of water environmental capacity and pollution load reduction using QUAL2K for water environmental management. International Journal of Environmental Research & Public Health, 2012, 9(12), pp.4504-4521.

[152] Zhang S J, Peng W Q. Water temperature structure and influencing factors in Ertan Reservoir, Shuili Xuebao/Journal of Hydraulic Engineering, 2009, 40(10), pp. 1254–1258.

[153] Zhang Y F, Wang Y L, Wang J. EFDC overview and application analysis. Environmental Impact Assessment, 2015, 37(3), pp.70-72. ality model, because they focused on different fields or app

[154] Zheleznyak M J, Demchenko R I, Khursin S L, Kuzmenko Y I, Tkalich P V, Vitiuk N Y. Mathematical modeling of radionuclide dispersion in the Pripyat-Dnieper aquatic system after the Chernobyl accident, Science of the Total Environment, 1992, 112(1), pp. 89–114.

[155] Zheng Z M, Xia B, Zhang J L. Research and practice on integration of GIS and water quality model. Environmental Science and Technology, 2005, (28), pp.10-12. (In Chinese)

Appendix I

==

1) Taking quantity of model simulation schemes currently executed by each execution node as the judgment basis, we find the optimal node with least execution number.

==

```
Public bool ExecuteEFDC(ServerDataSourcesLib.TaskNodes taskNodes,ServerConnector pConnector,
DataPackage schemaDataPack,string schemaFilePath)
{     if (taskNodes.Count <= 0) return false;
      ServerDataSourcesLib.TaskNodes lastNodes = new ServerDataSourcesLib.TaskNodes();
      DataPackageCreator dataCreator = new DataPackageCreator();
      dataCreator.PackageType = 0;
      DataPackage dataPack = dataCreator.CreateDataPackage("EFDCRun");
      for (int k = 0; k < taskNodes.Count; k++)
      {     ServerDataSourcesLib.TaskNodes tempNodes = new ServerDataSourcesLib.TaskNodes();
              ServerDataSourcesLib.TaskNode tempNode = taskNodes.get_Item(k);
              tempNodes.Add(tempNode);
              DataPackage rdataPack = pConnector.SendMessageByNodes(dataPack, tempNodes);
              if (rdataPack == null) continue;
              int count = rdataPack.GetNextInt();
              tempNode.Reserved = count;
              lastNodes.Add(tempNode);
           }
           if (lastNodes.Count <= 0) return false;
           lastNodes.Reorder(TaskNodePriorty.tnpReserved);
       ServerDataSourcesLib.TaskNodes optimalNodes = new ServerDataSourcesLib.TaskNodes();
       optimalNodes.Add(lastNodes.get_Item(0));
      DataPackage rdp = pConnector.SendMessageByNodes(schemaDataPack, optimalNodes);
       if (rdp == null)
       {
              return false;
       }
      byte result = rdp.GetNextByte();
      if (result == 1)
      {
               lastNodes.RemoveAt(0);
                 ExecuteEFDC(taskNodes, pConnector, schemaDataPack, schemaFilePath);
       }
    else if (result == 2)
    {     return true;
    }     else if (result == 10)
    {       lastNodes.RemoveAt(0);
           ExecuteEFDC(taskNodes, pConnector, schemaDataPack, schemaFilePath);
     }
       else if (result == 11)
```

```
{            lastNodes.RemoveAt(0);
             ExecuteEFDC(taskNodes, pConnector, schemaDataPack, schemaFilePath);
 }
 else if (result == 12)
 {      lastNodes.RemoveAt(0);
        ExecuteEFDC(taskNodes, pConnector, schemaDataPack, schemaFilePath);
  }
 else if (result == 13)
 {       lastNodes.RemoveAt(0);
         ExecuteEFDC(taskNodes, pConnector, schemaDataPack, schemaFilePath);}
  return false;
}
```

==

2) Inquire local scheme files through scheme ID.

==

```
DatabaseConnection conn = sysVar.CreateDatabaseConnection();
string schemaPath = Application.StartupPath + "\\files\\" + localSchemas.DefaultDirectory + "\\" + id + "." +
localSchemas.DefaultFileExa;
if (!File.Exists(schemaPath))
{      MessageBox.Show("The scheme doesn’t exist");
       return;
}
```

==

3) Read data of local scheme and transmit to the optimal execution node.

==

```
WaterOrinaryMonitorSchema schema = new WaterOrinaryMonitorSchema(sysVar); FileStream fstream = new
FileStream(schemaPath, FileMode.Open, FileAccess.Read);
BinaryReader breader = new BinaryReader(fstream);
byte[] datas = new byte[fstream.Length];
fstream.Seek(0, SeekOrigin.Begin);
breader.Read(datas, 0, datas.Length);
DataPackageCreator dataCreator = new DataPackageCreator();
dataCreator.PackageType = 1;
dataCreator.AddString(id);
dataCreator.AddBytes(datas);
DataPackage dataPack = dataCreator.CreateDataPackage("EFDCRun");
ServerConnector pConnector = new ServerConnector();
pConnector.ServerIP = conn.ServerIP;
pConnector.ServerPort = conn.ServerPort;
ServerDataSourcesLib.TaskNodes taskNodes = pConnector.GetBestNodes(TaskNodePriorty.tnpReserved, 5);
if (taskNodes.Count <= 0)
{    MessageBox.Show("The node server does’t exist");
     return;
```

```
  }
string schemaFilePath = "EFDCModels\\ordinary\\" + id + "." + localSchemas.DefaultFileExa;
================================================================================
```

4) After the execution node receives the scheme, it will store into corresponding catalogue and conduct unique identification by naming files with scheme ID.

```
String id = package.GetNextString();
Array ay = package.GetNextBytes();
byte[] datas = (byte[])ay;
//write in file
string filePath = parameters.WorkDirectory + "\\EFDCModels\\ordinary\\" + id + ".osh";
if (!Directory.Exists(Path.GetDirectoryName(filePath)))
Directory.CreateDirectory(Path.GetDirectoryName(filePath));
if (!File.Exists(filePath))
{       FileStream fileStream = new FileStream(filePath, FileMode.Create, FileAccess.Write);
        fileStream.Write(datas, 0, datas.Length);
        fileStream.Close();
  }
================================================================================
```

5) This execution node will make one copy of Mudan River water quality model files from template and conduct unique identification by naming file folders with scheme ID.

```
================================================================================
string sourceDir = parameters.WorkDirectory + "\\EFDCModels\\template";
if (!Directory.Exists(sourceDir))
Directory.CreateDirectory(sourceDir);
string targetDir = parameters.WorkDirectory + "\\EFDCModels\\ordinary\\" + id;
if (!CopyDirectory(sourceDir, targetDir))
   {       pCreator.AddByte(11); //Server copy the scheme failed!
           return pCreator;
     }
================================================================================
```

6) This execution node reads scheme data and substitute it for copied Mudan River water quality model files.

```
================================================================================

Schema.SamplePointsCol.PickTimeData(schema.FromTime, schema.ToTime, 24);
WaterMonitorProfileFiles profielFiles = new WaterMonitorProfileFiles(schema);
if (!profielFiles.OutputFiles(targetDir))
{     pCreator.AddByte(12);//Server replace the scheme failed.
      return pCreator;
}
```

```
EFDCInteraction pInteraction = new EFDCInteraction(schema);
try
  {     if (!pInteraction.SaveAll(targetDir))
        {     string error = pInteraction.ErrorInfo;
              pCreator.AddByte(12);
                return pCreator;
            }
     }
          catch (Exception ex)
            {     pCreator.AddByte(12);
                   return pCreator;
            }
```

==

7) This execution node tries to operate this Mudan River water quality model after replacement in multithreading.

==

```
public int ExecuteEFDC(string pathName, int waitMiniSeconds)
{     lock (m_threaddata)
        {     for (int k = 0; k < THREAD_COUNT; k++)
                {
                        if (m_threaddata[k].pathName == pathName) return 0;
                }
                int dwRet = WaitHandle.WaitAny(m_signals, waitMiniSeconds);
                if (dwRet == 258) return 1;
                m_threaddata[dwRet].pathName = pathName;
                m_events[dwRet].Set();
        }
         return 2;
}
```

==

8) If the operation is successful, then the route where this scheme is located will be registered into distributed file system.

==

```
if (rt == 2)

{       string fileDir = "EFDCModels\\ordinary";
          DataSourcesLib.FileDesp desp = pFileSys.FindDirectory(fileDir);
        if (desp.Name == "")
        {
                pFileSys.ServerConnector.CreateFileSysConnection().AddDirectoryCascade("EFDCModels");
                pFileSys.ServerConnector.CreateFileSysConnection().AddDirectoryCascade("EFDCModels\\ordinary"
```

```
            );
        }
        fileDir = fileDir + "\\" + id + ".osh";
        pFileSys.MountFile(fileDir);
    }
```
==

9) If among previous step 3) to 7), error information is returned, the system will eliminate this node and return to step 1) until the scheme is successfully operated.

10) After execution node operation is finished in water quality forecasting model, this execution node will register completed logo files completed in model operation into distributed file system.

==

```
  //Register the scheme to DFS
ServerFileSys pFileSys = new ServerFileSys();
IServerConnector pConnector = pFileSys.ServerConnector;
pConnector.ServerIP = parameters.MainServerIP;
pConnector.ServerPort = parameters.CloudServerPort;
string[] arrStr = dir.Split('\\');
string id = arrStr[arrStr.Length -1];
string idPath = "EFDCModels\\ordinary\\" + id;
DataSourcesLib.FileDesp desp = pFileSys.FindDirectory(idPath);
if (desp.Name == "")
{pFileSys.ServerConnector.CreateFileSysConnection().AddDirectoryCascade("EFDCModels");
pFileSys.ServerConnector.CreateFileSysConnection().AddDirectoryCascade("EFDCModels\\ordinary");
pFileSys.ServerConnector.CreateFileSysConnection().AddDirectoryCascade(idPath);
pFileSys.ServerConnector.CreateFileSysConnection().AddDirectoryCascade(idPath + "\\#output");
}
 pFileSys.MountFile(idPath + "\\#output\\TIME.LOG");
 pFileSys.MountFile(idPath + "\\DYEDMPF.ASC");
 break;
```
==

11) Download simulation results, through the distributed file system, we will inquire which route of which elastic execution node the simulation data of the scheme is located, and then we conduct corresponding downloading.

==

```
DatabaseConnection conn = sysVar.CreateDatabaseConnection();
ServerConnector pConnector = new ServerConnector();
pConnector.ServerIP = conn.ServerIP;
pConnector.ServerPort = conn.ServerPort;
```

```
IServerFileConnector fConnector = (IServerFileConnector)pConnector;
string resultPath = "EFDCModels\\ordinary\\" + id + "\\DYEDMPF.ASC";
ServerDataSourcesLib.TaskNodes nodes = fConnector.GetAllNodes(resultPath, ServerFileType.sftFile);
if (nodes.Count <= 0)
{
    MessageBox.Show("Node doesn’t exist for store simulation result.");
    return;
}
ServerDataSourcesLib.TaskNode node = nodes.get_Item(0);
ClientSocket socket = new ClientSocket();
int iip = node.ServerIP;
NetServer netServer = new NetServer();
string ip = netServer.ConvertNumberToIP(iip);
socket.ServerIP =  ip;
socket.ServerPort = node.ServerPort;
DataPackageCreator packageCreator = new DataPackageCreator();
packageCreator.PackageType = 3;
packageCreator.AddString(id);
DataPackage packge = packageCreator.CreateDataPackage("EFDCRun");
if (!socket.Connect())
{   MessageBox.Show("failure to connect the node server");
    return;
}
DataPackage rebackPackge = socket.SendMessage(packge);
Array dy = rebackPackge.GetNextBytes();
byte[] results = (byte[])dy;
MemoryStream ms = new MemoryStream();
ms.Write(results, 0, results.Length);
ms.Seek(0, SeekOrigin.Begin);
StreamReader sr = new StreamReader(ms);
//read data
WaterOrinaryMonitorSchema schema = new WaterOrinaryMonitorSchema(sysVar);
string localPath = Application.StartupPath + "\\files\\" + localSchemas.DefaultDirectory + "\\" + id + "." +
localSchemas.DefaultFileExa;
if (!schema.ReadFromBinary(sysVar, localPath))
{   MessageBox.Show("Failure to read the scheme file.");
    return;
}
ConcentResults crs = new ConcentResults(schema);
crs.ReadResultsFromStreamReader(sr);
sr.Close();
ms.Close();
```

==

12) After simulation results files are downloaded, we’ll conduct reading in GIS client side and conduct spatial-temporal dynamic display.

```
==============================================================================/
//add grids

  private void AddGridAreaLayer()
        {
            WaterMonitorProfileConcentResultFeatureWorkspaceFactory pFc = new WaterMonitorProfileConcentResultFeatureWorkspaceFactory();
            WaterMonitorProfileConcentResultFeatureWorkspace newwork = pFc.CreateWorkspace();
            WaterOrdinaryMonitorProfileConcentResultFeatureClass newfc = new WaterOrdinaryMonitorProfileConcentResultFeatureClass(new WaterMonitorConcentResultFeatureDataset(newwork, "daily simulation", schema));
            newfc.Results = results;
            newfc.SysVar = pContainer.MainFrame.SystemVariables;
            FeatureLayer pLayer = new FeatureLayer();
            pLayer.FeatureClass = newfc;
            pLayer.UserLabel = "Concentration in grid";
            pLayer.Transparent = 50;
            updateLegend(pLayer, 0);
            BaseObjectsLib.TreePath pPath = new BaseObjectsLib.TreePath();
            pPath.AddChildNode(0);
            BasicMap pMap = mapCol.Maps.ActiveMap;
            pMap.InsertLayer(pPath, pLayer);
            pMap.ZoomToLayer(pLayer);
        }

//update legend
  private void updateLegend(FeatureLayer lLayer, int state)
        {   if (state == 0)
            {   GradualColors pgc = new GradualColors();
                pgc.AddColorBreak(0, 0, 255);
                pgc.AddColorBreak(0, 255, 0);
                pgc.AddColorBreak(255, 255, 0);
                pgc.AddColorBreak(255, 0, 0);
                ClassBreakLegend pcb = new ClassBreakLegend();
                pcb.field = "concentration";
                float intervalValue = results.MaxGrid.value / 29;
                for (int k = 0; k < 29; k++)
                { pcb.AddBreak(intervalValue * k);
                }
                for (int k = 0; k < pcb.Count; k++)
                {   SimplePolygonSymbol pSymbol = new SimplePolygonSymbol();
                    SimpleLineSymbol lineSymbol = new SimpleLineSymbol();
                    RGBColor color = new RGBColor();
                    color.SetNullColor();
                    lineSymbol.color = color;
                    pSymbol.Outline = lineSymbol;
```

```
                pcb.Symbol[k] = pSymbol;
            }
            pcb.SetSymbolsColor(pgc);
            pcb.InitialLabelsWithRegionString();
            pcb.Description = schema.WaterQualityDecayItem.WaterQualityItem.Name + "/（" +
schema.WaterQualityDecayItem.WaterQualityItem.Unit+ "）";
            lLayer.SetLegend(pcb);
        }
        else
        {   UniqueLegend uniqueLegend = new UniqueLegend();
            uniqueLegend.AddField("grades");
            uniqueLegend.AddUnique("Ⅰ", "Ⅰ");
            uniqueLegend.AddUnique("Ⅱ", "Ⅱ");
            uniqueLegend.AddUnique("Ⅲ", "Ⅲ");
            uniqueLegend.AddUnique("Ⅳ", "Ⅳ");
            uniqueLegend.AddUnique("Ⅴ", "Ⅴ");
            uniqueLegend.AddUnique("+Ⅴ", "+Ⅴ");
            RGBColor[] colors = new RGBColor[6];
            RGBColor color = new RGBColor();
            color.PutRGB(0, 0, 255);
            colors[0] = color;
            color = new RGBColor();
            color.PutRGB(0, 255, 0);
            colors[1] = color;
            color = new RGBColor();
            color.PutRGB(0, 255, 255);
            colors[2] = color;
            color = new RGBColor();
            color.PutRGB(255, 0, 255);
            colors[3] = color;
            color = new RGBColor();
            color.PutRGB(255, 255, 0);
            colors[4] = color;
            color = new RGBColor();
            color.PutRGB(255, 0, 0);
            colors[5] = color;
            for (int k = 0; k < uniqueLegend.Count; k++)
            {
                SimplePolygonSymbol pSymbol = new SimplePolygonSymbol();
                SimpleLineSymbol lineSymbol = new SimpleLineSymbol();
                RGBColor nullColor = new RGBColor();
                nullColor.SetNullColor();
                lineSymbol.color = nullColor;
                pSymbol.Outline = lineSymbol;
                pSymbol.color = colors[k];
                uniqueLegend.set_Symbol(k, pSymbol);
```

```
                }
            uniqueLegend.Description = schema.WaterQualityDecayItem.WaterQualityItem.Name + "/（" +
                        schema.WaterQualityDecayItem.WaterQualityItem.Unit + "）";
        lLayer.SetLegend(uniqueLegend);
            }
        }
//update map layers
  private   void RefreshLayer(string userLabel)
        {     BasicMap pMap = mapCol.Maps.ActiveMap;
              FeatureLayer lLayer = (FeatureLayer)ThematicLayerManager.FindLayer(mapCol.Maps.ActiveMap, userLabel);
              if (lLayer == null) return;
              IFeatureClass lfc = lLayer.FeatureClass;
              WaterOrdinaryMonitorProfileConcentResultFeatureClass wmfc = lfc as WaterOrdinaryMonitorProfileConcentResultFeatureClass;
              wmfc.CurrentFrame = curFrame;
              if (lLayer != null)
              {     ILayer pLayer = lLayer as ILayer;
                    pMap.RefreshLayer(lLayer);
              }
        }
//the first function to call FeatureClass of grid
  public class WaterOrdinaryMonitorProfileConcentResultFeatureClass : IFeatureClass
    {
        protected WaterOrinaryMonitorSchema schema;
        protected SpatialReference psp = new SpatialReference();
        protected string name = "daily simulation";
        protected SelectionSet pSelectionSet = new SelectionSet();
        protected ITableDesc pDesc;
        protected WaterMonitorConcentResultFeatureDataset pDataset;
        protected FeatureClassTool pTool = new FeatureClassTool();
        ConcentResults results;
        SystemVariables sysVar;
        protected EnvelopeDef fullExtent = null;
        int currentFrame = 0; //the 0 frame means initial concention，other frames mean simulation result of concentration
        byte[] cellTypes; //grid’s properties
        float[] concentTypes;   //initial concentration
        public WaterOrdinaryMonitorProfileConcentResultFeatureClass(WaterMonitorConcentResultFeatureDataset pDataset)
        {
              this.pDataset = pDataset;
              this.schema = pDataset.Schema as WaterOrinaryMonitorSchema;
              psp = schema.SpatialReference;
              TableDescCreator pCreator = new TableDescCreator();
              pCreator.AddField("current time", VectorFieldType.vftDateTime, 10);
```

```
            pCreator.AddField("concentration", VectorFieldType.vftReal, 10, 2);
            pCreator.AddField("grid", VectorFieldType.vftString, 10);
            pCreator.AddField("code", VectorFieldType.vftString, 20);
            pDesc = pCreator.CreateTableDesc();
            fullExtent = schema.ProfileGrid.GetFullExtent();
            pTool.FeatureClass = this;
            cellTypes = ((WaterMonitorProfileCellTypeProperty)schema.ProfileGridEntity["grid's
properties"]).RawData;
            concentTypes =
((WaterMonitorProfileConcentraProperty)schema.ProfileGridEntity["concentration"]).rawData;
        }
        public SystemVariables SysVar
        {   set
            { sysVar = value;
            }
        }
        public ConcentResults Results
        {   set
            {   results = value;
                currentFrame = 0;
            }
        }
        public int CurrentFrame
        {   set
            {
                int cf = value;
                if (cf < 0 || cf > results.Count) return;
                currentFrame = cf;
            }
        }
        public string ClassName
        {   get { return "daily simulation feature class"; }
        }
        public int FeatureCount
        {
            get
            {   return ((IWaterMonitorProfileGridArea)schema.ProfileGrid).AreasCount;
            }
        }
        public IFeatureDataset FeatureDataset
        {   get { return pDataset; }
        }
        public int FindFeatureID(int FeatureID)
        {   return FeatureID;
        }
        public IGeometry GetFeature(int RecordIndex)
```

```
{       if (cellTypes[RecordIndex] != 5) return null;
        return schema.ProfileGridEntity.GetGeometry(RecordIndex);
}
public bool GetFeatureEnvelope(int RecordIndex, EnvelopeDef pEnv)
{       if (schema == null) return false;
        if (cellTypes[RecordIndex] != 5) return false;
        schema.ProfileGridEntity.GetFeatureEnvelope(RecordIndex, ref pEnv);
          return true;
}
public int GetFeatureID(int RecordIndex)
{       return RecordIndex;
}
public ConcentResult GetCurrentResult()
{       return results.GetResult(currentFrame - 1);
}
private int GetAreaWaterConcentGrade(float value)
{       WaterQualityItem item = schema.WaterQualityDecayItem.WaterQualityItem;
        if (item == null) return -1;
        int vstate = item.ValueState;
        if (vstate > 1) return -1;
        if (vstate == 0)
        {       for (int j = 0; j < 5; j++)
                {
                        if (value <= item.Standards[j]) return j;
                }
                return 5;
        }
          if (vstate == 1)
          {     for (int j = 0; j < 5; j++)
                {
                        if (value >= item.Standards[j]) return j;
                }
                return 5;
          }
        return    -1;
}
private string GetAreaWaterConcentGradeAsString(float value)
{       string sts = string.Empty;
        int igrade = GetAreaWaterConcentGrade(value);
        switch (igrade)
        {       case 0:
                case 1:
                        sts = "excellent";
                        break;
                case 2:
                        sts = "good";
```

```
                break;
            case 3:
                sts = "mild pollution";
                break;
            case 4:
                sts = " moderate pollution";
                break;
            case 5:
                sts = "serious pollution";
                break;
            default:
                sts = "no data";
                break;
        }
        return sts;
    }
```

Gula TANG

Research on Distributed Warning System of Water Quality in Mudan River Based on EFDC and GIS

Résumé

Le système de simulation et d'avis précoce d'alerte est un outil puissant pour la surveillance de la qualité de l'eau de la rivière. Mudan, une rivière importante dans les régions froides du nord-est de la Chine et qui se jête finalement dans la rivière de l'Amour en Russie. Ainsi la qualité de l'eau dans la rivière Mudan est une préoccupation importante non seulement au niveau local et régional, mais aussi au niveau international. L'objectif de cette étude est de créer un système de simulation et d'avis précoce d'alerte pour que la distribution spatio-temporelle de la qualité de l'eau durant les périodes de couverture glaciaire et d'eaux libres soit simulée et visualisée précisément et afin que l'on puisse appréhender la variation spatiale de polluants sur le cours de rivière. La thèse est structurée en 7 chapitres. Dans le premier chapitre nous décrivons le contexte de l'étude et faisons un état de lieu des recherches actuelles. Dans le chapitre II, la comparaison des modèles principaux disponibles pour l'évaluation de la qualité de l'eau est réaliser ainsi que le choix du meilleur modèle comme base pour créer le système de modélisation. Dans le chapitre III, la construction du modèle, les conditions limites requises et les paramètres pour le modèle ont été vérifiés et étalonnés. Une procédure de simulation distribuée est conçue dans le chapitre IV pour améliorer l'efficacité de la simulation. Le chapitre V concerne la programmation et la réalisation la de simulation distribuée et le chapitre VI les techniques fondamentales pour mettre en œuvre le système. Le chapitre VII est la conclusion. Il y a trois points innovants dans ce travail: un modèle bidimensionnel de dynamique de fluides de l'environnement pour la rivière Mudan, une méthode efficace du calcul distribué et un prototype de système de simulation et d'avis précoce d'alerte qui peuvent largement améliorer la capacité de surveillance et de gestion de la qualité de l'eau de la rivière Mudan ou d'autres rivières similaires.

Mots clés: préavis d'alerte; modèle hydrodynamique; qualité de l'eau; simulation bidimensionnelle; calcul distribué; analyse de scénarios; SIG

Résumé en anglais

Simulation and Early Warning System (SEWS) is a powerful tool for river water quality monitoring. Mudan River, an important river in northeastern cold regions of China, can run out of China into Russia. Thus, the water quality of Mudan River is highly concerned not only locally and regionally but also internationally. Objective of this study is to establish an excellent SEWS of water quality so that the spatio-temporal distribution of water quality in both open-water and ice-covered periods can be accurately simulated and visualized to understand the spatial variation of pollutants along the river course. The dissertation is structured into 7 chapters, chapter I outlines the background of the study and reviews the current progress. Chapter II compares the main available models for evaluating river water quality so that a better model can be selected as the basis to establish a modeling system for Mudan River. Chapter III establishes the model, the required boundary conditions and parameters for the model were verified and calibrated. Chapter IV, a distributed simulation procedure was designed to increase the simulation efficiency. Chapter V discusses more about the programing and operational issues of the distributed simulation. Chapter VI is about the core techniques to implement the system. Chapter VII is the conclusion of the study to summarize the key points and innovations of the study. The study has the following three points as innovation: a two-dimensional environmental fluid dynamics model for Mudan River, an efficient distributed model computational method and a prototype of SEWS, which can greatly improve the capability of monitoring and management of water quality in Mudan River and other similar rivers.

Keywords: early warning; hydrodynamic model; water quality; two-dimensional simulation; distributed computation; scenario analysis; GIS