UNIVERSITÉ CHEIKH ANTA DIOP DE DAKAR

ECOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE

FACULTÉ DES SCIENCES ET TECHNIQUES

Année : 2019 /2020 N° d'ordre : 168

THÈSE DE DOCTORAT UNIQUE

Mention : Mathématiques et modélisation

Spécialité : Analyse, statistiques et applications.

PRESENTEE PAR : **TRAORE ABOUBAKARI**

TITRE

Contributions numériques pour la résolution de certains problèmes environnementaux et épidémiologiques.

Soutenue publiquement le 26/12/2020 devant le jury composé de :

Hamidou DATHE	**Professeur Titulaire**, Université Cheikh Anta Diop, Dakar (Sénégal),	Président
Abdou NJIFENJOU	**Professeur Titulaire**, Université de Yaoundé 1, (Cameroun),	Rapporteur
Blaise SOME	**Professeur Titulaire**, Université de Ouagadougou, (Burkina Faso),	Rapporteur
Roger Marcelin FAYE	**Professeur Titulaire**, Université Amadou Mahtar Mbow (Sénégal),	Examinateur
Alassane SY	**Maître de conférences**, Université Alioune Diop, Bambey (Sénégal),	Examinateur
Benjamin MAMPASSI	**Professeur Titulaire**, Université Cheikh Anta Diop, Dakar (Sénégal),	Directeur

Résumé

Ce travail est une contribution à la modélisation et simulation numérique d'une part aux problèmes liés à la pollution et l'assèchement des eaux de surface, d'autre part aux problèmes de propagation d'une maladie dans une population. Le but recherché dans cette thèse est de mettre en place un outil d'aide à la décision dans la recherche des solutions à certains problèmes cités plus haut. Pour atteindre cet objectif, nous avons développé des modèles mathématiques et des schémas numériques adaptés aussi bien dans le cas de la pollution qu'en épidémiologie.

Par ailleurs, dans le cas de la pollution des eaux, le traitement numérique des domaines d'étude, considérés comme ayant des circonférences non régulières, nous a amené à faire recours au couplage de techniques déjà existantes pour le maillage du domaine. Cette approche a fortement amélioré les résultats obtenus.

De plus en épidémiologie, le paramétrage de l'échelle de temps des fonctions vitales a conduit à des équations singulièrement perturbées. Alors Nous avons fait une analyse asymptotique du modèle obtenu et évalué les erreurs commises.

A chacune des catégories de problèmes que nous avons présentées, nous avons appliqué les solutions développées à un exemple concret et les résultats de ces travaux ont fait l'objet de publications scientifiques que nous avons jointes en annexe de cette thèse.

Les fruits de ces travaux de recherche ont été consignés dans une plateforme que nous avons développée sur MATLAB appelée interface graphique de résolution des problèmes en environnement et en épidémiologie. Cette application constitue la base de notre progiciel. Quelques résultats de cette application ont été présentés au chapitre quatre de cette thèse.

Mots Clés : *Systèmes distribués à données incomplètes- Propagation de pollutions dans les fluides – Méthodes pseudo spectrale- Méthode de Heun- Méthodes d'éléments finis colocaux moindres carrés- Simulation numérique*-Methode asymptotique.

Dédicace

Je dédie cette thèse à mes deux enfants : Seydina-Aly et Abdul-Razak TRAORE et leur mère KA epse TRAORE FATOUMATA.

Remerciements

Je tiens à remercier très sincèrement le Professeur **Benjamin MAMPASSI** pour sa patience dans l'encadrement de cette deuxième thèse. Auprès de lui je continue d'apprendre les valeurs humaines que son humilité et le sens du travail bien fait.

J'exprime ma gratitude à Monsieur **Blaise SOME**, Professeur titulaire à l'université de Ouagadougou (Burkina Faso) pour avoir accepté d'examiner cette thèse et pour ses nombreux conseils dans la recherche lors des différents colloques que nous avions passés ensemble.

Je tiens aussi à remercier Monsieur **Abdou NJIFENJOU**, professeur titulaire à l'université de Yaoudé 1 (Cameroun) d'avoir accepté d'examiner cette thèse.

Je remercie les membres du jury :

- Monsieur **Hamidou DATHE**, Professeur titulaure au département de mathématiques et informatique d'avoir accepté de présider ce jury.
- Monsieur **Roger Marcelin FAYE**, Professeur titulaure à l'université Amadou Moctar Mbow pour sa participation au jury.
- Monsieur **Alassane SY**, maître de conférences à l'université Alioune Diop de Bambey pour sa participation au jury.

J'exprime ma gratitude **aux collègues de l'équipe d'analyse numérique** auprès du Professeur MAMPASSI et tout particulièrement à **Dr. Bassirou DIA** pour sa disponibilité et son aide pendant la constitution des dossiers de soutenance.

J'adresse mes vifs remerciements à ma belle famille : **la famille KA à M'bour**, **la famille BOUT à SALLY**,**la famille Diallo à Rufisque** et la **famille DIALLO à Dakar** pour leurs accueils très chaleureux lors de mes séjours au Sénégal dans la préparation de cette thèse.

je ne saurai terminé sans avoir une pensée pour mon **Papa Sinaly TRAORE**. Que le Seigneur lui accorde son paradis. Merci beaucoup **Maman Fatoumata OUATTARA** pour tes nombreuses prières, Que le Seigneur t'accorde longue vie.

Liste des tableaux

Table des figures

Table des matières

Introduction

Cette thèse se présente comme une synthèse des travaux de recherches réalisés après notre thèse de troisième cycle [48] en 2008. Elle s'inscrit dans la réalisation des perspectives annoncées dans cette dernière, celles de mettre en place un progiciel, un outil d'aide à la décision, dans la recherche des solutions à certains problèmes liés à l'environnement et à la santé. Particulièrement, nous avons travaillé d'une part, sur la modélisation mathématique de problèmes de pollution des eaux en surface et d'épidémiologie, et d'autre part, sur le développement d'algorithmes de calcul.

Dans ce premier chapitre, nous commencerons par le contexte et la problématique des thématiques de ce travail, ensuite nous poserons les questions majeures de recherche qui ont motivé ce travail, après nous présenterons les contributions majeurs obtenues pendant ces recherches et nous terminerons par la démarche adoptée pour présenter cette thèse.

Contexte et problématique

De nos jours, les problèmes liés à l'environnement sont au cœur de toutes les grandes rencontres internationales des hommes politiques jusqu'aux organisations non gouvernementales. La pollution de l'environnement, limitée il y a quelques dizaines d'années dans les pays industrialisés, est devenue un problème mondial et concerne de plus en plus les pays en voie de développement, pourtant classés parmi

les pays les moins pollueurs de la planète. Les perturbations environnementales dues aux activités anthropogéniques telles que l'usage des pesticides dans l'agriculture et le rejet des déchets industriels et domestiques, se multiplient dans de nombreuses régions du globe et entrainent diverses pollutions des eaux tant souterraines que superficielles [53]. La dégradation de l'environnement est due également au changement climatique qui se manifeste par un réchauffement global, et donc une augmentation de la température des eaux de surface (cours d'eau, lacs, mers). Ces pics de chaleurs ont un impact sur le niveau des nappes phréatiques et des rivières. Il y a moins d'eau disponible, donc des polluants plus concentrés.

Selon des rapports de l'OMS [42] et de l'UNESCO [50], les problèmes liés à la pollution de l'environnement entraineront une diminution de la production vivrière dans de nombreuses régions parmi les plus démunies. Cette diminution atteindra jusqu'à 50% d'ici 2020 dans certains pays africains. En conséquence, Il en résultera une prévalence accrue de la malnutrition et de la dénutrition, actuellement à l'origine de 3,1 millions de décès par an. Aussi, selon l'OMS, le changement climatique pourrait entraîner environ 250 000 décès supplémentaires par an entre 2030 et 2050 : 38 000 dus à l'exposition à la chaleur des personnes âgées, 48 000 dus à la diarrhée, 60 000 dus au paludisme, et 95 000 dus à la sous-alimentation des enfants .

Des maladies hydriques où environ 1.7 millards de personnes en souffrent chaque année, 50% des décès infantiles étant enregistrés en Afrique.

Face à ce grand défis, les politiques et les scientifiques sont tous mobilisés pour apporter des réponses aux problèmes de la pollution de l'environnement. Dans la suite nous présenterons des questions de recherches liées à la pollution des eaux et en épidémiologie auxquelles nous avons proposé des réponses.

Problèmes de recherches

Les thématiques auxquelles nous sommes confrontées dans cette thèse sont liées à la problématique de la pollution des eaux de surface et celle de la propagation d'une épidémie, même si l'un pourrait être considéré comme une conséquence de l'autre. La conception d'un modèle mathématique décrivant ces phénomènes ne saurait mieux décrire aussi précise que possible la réalité sans une prise en compte de toutes les propriétés les caractérisant.

Ainsi, la question d'identification des paramètres du modèle mathématique à partir de données réelles est fondamentale dans le processus de modélisation. On parle de problème inverse. On trouve une littérature assez fournie sur la résolution des problèmes inverses : voir par exemple les références [4] et [34]. On peut se référer également aux travaux de l'équipe de J. P. KERNEVEZ pour le traitement numérique de problèmes d'identification de pollutions dans les systèmes distribués ([30] et [37]), la détection de pollution dans un aquifer [14], la détermination de paramètres manquants dans un lac et la recherche de pollution dans une rivière ([5] et [30]). La question fondamentale dans l'étude de la pollution des eaux de surface à laquelle nous nous intéressons est : Est il possible d'identifier de façon complète les fonctions ou paramètres inconnus à partir des données prélevées en prenant en compte les erreurs qui peuvent subvenir sur les mesures ?

La modélisation en épidémiologie a beaucoup progressé ces trentes dernières années : des modèles différentiels ordinaires développés à partir des modèles compartimentaux ([6], [19] et [25]) aux systèmes d'équations aux dérivées partielles dépendant du temps et souvent de l'âge des individus, de la population totale et du sexe des individus. Pour une étude intensive sur la modélisation et la théorie sur l'existence des modèles structurés en âge nous invitons le lecteur à voir les travaux de Iannelli ([27], [28]) et pour ce qui est du comportement numérique des problèmes (voir [52]) . D'excellents travaux ont été publiés dans le cadre du développement et de l'analyse de schémas numériques pour les modèles dépendant de l'âge. L'on pourrait consulter

les travaux suivants : [3], [7] et [35]. Dans cette thématique, le problème auquel nous nous intéressons est d'une part la mise en place d'un schéma numérique pour un domaine d'âge illimité et d'autre part la mise en place d'un modèle mathématique de l'évolution de l'hépatite B en prenant en compte de nouvelles données plus réalistes.

Cependant, beaucoup reste à faire quant au traitement numérique de ces modèles. Les outils de simulation numérique disponibles sont encore perfectibles pour répondre aux nombreuses attentes des problèmes posés.

Dans la suite, nous proposons des solutions apportées aux questions citées plus haut.

Solutions proposées

Les solutions proposées se déclinent en modèles mathématiques développés et en schémas numériques conçus pour les résoudre.

Modèles mathématiques développés

Les modèles mathématiques développés dans cette thèse se resument en trois parties. Premièrement, l'étude de la pollution d'une étendue d'eau nous a conduit à nous intéresser à un moment donné au phénomène de l'assèchement d'une surface d'eau. En effet, un Lac ou une rivière soumis à une pollution constante et en grande quantité peut conduire au problème d'eutrophisation. Ce problème est une cause directe de la montée de la boue, de l'ensablement des fonds marrins voir de l'assèchement des cours d'eau. Dans un travail collaboratif [31], nous avons produit un modèle mathématique décrit par un système d'équations aux dérivées partielles de second ordre dont la variable $h(t, x)$ mesure la hauteur du niveau d'eau du fond du domaine à la surface du liquide à une position x. Ce modèle mathématique a été conçu et validé à partir de photographies satellitaires dans le cas du Lac Tchad sur la période de 1963 à 2001. Ce travail, publié en 2013 [31], même s'il est loin d'ex-

plorer toute la problématique de l'assèchement d'un Lac, a permit de développer un modèle mathématique sur la base de données photographiques satellitaires.

Ensuite, en épidemiologie, notre étude a porté sur des populations en présence d'une maladie avec des forces de contaminations variées. Dans cette seconde phase, notre attention s'est portée sur la maladie de l'hépatite B. Il est bien connu que cette maladie fait l'objet de beaucoup de recherche dans le monde scientifique car n'ayant pas de traitement de nos jours. Nous avons apporté une contribution au modèle développé dans [55] en considérant d'autres réalités de la maladie notamment au niveau de la fonction d'infection pour un meilleur suivi de la propagation de la maladie. Ces contributions ont aboutit à la publication de l'article [36] en 2017.

Enfin, nous avons développé un modèle sur la détection de pollution dans les eaux de surface. Nous avons reconsidéré une surface de l'eau ayant une géométrie complexe et cette eau est soumise à une pollution. La source de pollution reste inconnue et les conditions aux limites du domaimes sont partiellement connues. Tout comme le précedent travail, notre recherche a consisté à definir un modèle mathématique permettant de décrire la dynamique de la concentration de la pollution dans le fluide. Par ailleurs, il a consisté à déterminer la source de pollution et d'analyser l'évolution de la pollution dans le temps. Cette recherche a conduit à la publication d'un article scientifique [46] en 2017. Dans le même sens, nous avons développé un modèle dans le cas de la pollution d'une rivière qui est en cours de publication, [49] en 2020.

Tous ces modèles obtenus ont fait l'objet d'analyse et de simulation numérique. Dans la suite nous présenterons quelques unes de ces méthodes.

Schémas numériques développés

Deux méthodes numériques ont été explorées à travers nos travaux dans la mise en place de schémas numériques adéquats en vue de trouver des solutions approchées aux systèmes d'équations modélisant les problèmes cités dans le paragraphe précédent. Il s'agit des méthodes pseudo-spectrales et des méthodes de Runge Kutta. La

première méthode a été largement développée au cours de nos travaux de thèses de troisième cycle et même après. Elle nous a été fortement utile dans l'approximation des équations aux dérivées partielles définies sur des domaines spatiaux complexes en 2D. Elle a fait object de beaucoup de contributions en collaboration avec d'autres auteurs dans l'approximation des opérateurs de différentiation de second ordre sur des maillages non uniformes. Les résultats de ces travaux ont abouti à la publication d'articles scientiques : ([38]) en 2014 et ([39], [40] et [47]) en 2011 . Aussi, en se basant sur les travaux précédemment cités, nous avons développé un algorithme plus efficace pour la résolution d'un modèle mathématique décrivant un lac soumis à une pollution ([46] en 2017).

Les méthodes de Runge Kutta nous ont été utiles dans le traitement des modèles mathématiques en épidémiologie ayant plusieurs variables et une forte dépendance entre elles. Nous nous sommes intéressés à la résolution des équations dependant de l'âge avec une non-linéarité dans les fonctions de natalité, mortalité et d'infection. Nos travaux s'inspirent de l'article de [2] qui présente un algorithme général de résolution de tel système avec la méthode de Runge Kutta sur une durée de vie limitée. Dans nos travaux, nous supposons une durée de vie illimitée et avions amelioré le temps de convergence dans le cas de la méthode de Heun à l'ordre 2. Ceci a abouti à la publication de l'article [9].

GUI : Graphical User Interface

Dans cette dernière partie, nous présenterons une interface graphique (GUI) comme le début d'un progiciel pour certains problèmes numériques. Cette application est le fruit de différents programmes élaborés dans la résolution des équations aux dérivées partielles issues de la modélisation de problèmes de pollution ou épidémiologique.

Dans la suite de ce travail, le lecteur trouvera dans le chapitre 2, les travaux sur les modèles mathématiques et les méthodes numériques développés dans la résolution de problèmes de pollutions des eaux de surface et du traitement des domaines complexes. Le chapitre 3 est dévolu aux travaux sur la modélisation et simulation numérique de problèmes épidémiologies. Nous présenterons un modèle de GUI dans le chapitre 4 et enfin nous présenterons la conclusion générale au chapitre 5.

Chapitre 1

Modélisation et simulation numérique de problèmes liés à la pollution des eaux de surface [31, 38, 39, 40, 46, 47]

Ce chapitre est la synthèse de trois publications scientifiques, [31] (2013), [46](2017) et [47](2011)(voir les annexes 1, 5 et 2). Nous présenterons d'abord le contexte général sur les problématiques liées à la pollution des eaux de surface. Ensuite, nous étudierons les techniques de maillage des domaines à géométrie complexe qui aboutiront à la formulation des schémas numériques et enfin nous étudierons des problèmes tests.

1.1 Contexte et justification

Les eaux de surfaces se définissent comme une étendue d'eau dont la profondeur est négligeable par rapport à la largeur et à la longueur. On pourrait citer dans cette famille des lacs et des fleuves. Ainsi une eau de surface est soumise a une

pollution lorsqu'elle contient généralement des déchets chimiques comme le Nitrate et le Phosphore provenant des eaux usées, des industries, des eaux de ruissellement.

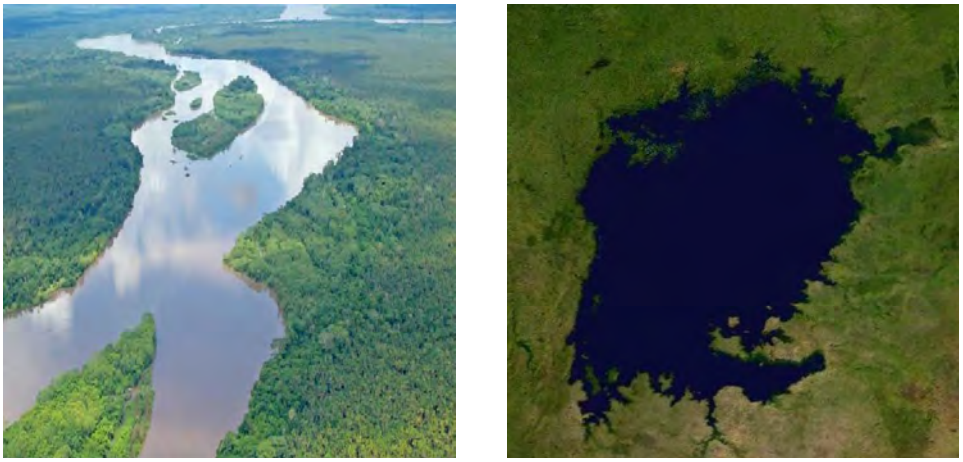
Ces polluants tuent les poissons et d'autres animaux aquatiques en redusisant le taux d'oxygène dissout dans l'eau. Ainsi, il existe une corrélation entre le taux d'oxygène dissout et le taux de pollution dans l'eau et c'est pourquoi la quantité de pollution est évaluée en fonction de la quantité d'oxygène dissout dont les polluants ont besoin pour leur reaction chimique et biologique. Cette quantité est estimée en BOD (Biologic Oxygen Demand) ou (demande biologique en Oxygène) et COD (Chemical Oxygen Demand) ou (demande chimique en Oxygène). Voir les travaux [18] et [30]. Nous supposerons dans ce paragraphe que $C(t, \mathbf{x})$ (Kg/m^3) désigne la concentration de la pollution dans une eau de surface mesurée en COD à l'instant t et repérée dans l'espace par $\mathbf{x}$.

1.2 Maillage du domaine

Dans cette section nous sommes concernés par des domaines de type Figure 1.1. A cet effet, nous présenterons un processus de maillage en deux dimensions qui utilise à la fois la méthode des éléments finis triangulaires et les points de collocation. Dans la suite de cette section, le paramètre NT contrôlera le nombre de triangles et le paramètre NP désignera le nombre de points collocaux dans chaque élément fini triangulaire.

1.2.1 Maillage en éléments finis triangulaires

Nous avons une variété de codes numériques disponibles pour générer des éléments finis triangulaires sur des figures à géométrie complexe ou irrégulière. L'on pourrait consulter [22] où l'auteur développe un algorithme pour le maillage des figures en dimension 2 ou 3 d'espace sans utiliser le critère de Delauney contrairement à [32]. Dans ce travail, nous présentons un algorithme qui inclut la dérivation



(a) (b)

Figure 1.1 – *(a)Le fleuve Amazone, (b)Le Lac Victoria*

de Delauney. Ce programme utilise des codes tels que : "*inpolygon*" et "*delaunay-Triangulation*" disponibles dans les outils de Matlab. Le principal avantage de notre programme est la rapidité et la légèreté permettant de créer facilement les éléments finis triangulaires. Quelques exemples de maillages sont présentés dans Figure 1.2.

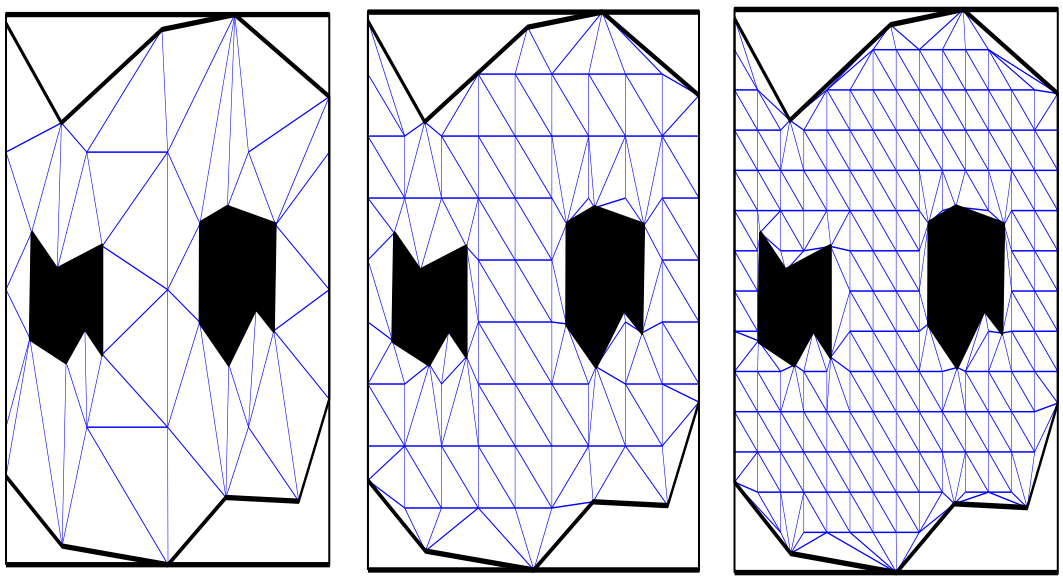


Figure 1.2 – *De gauche à droite, le maillage en éléments finis triangulaires pour respectivement* $NT = 5$, 10 *et* 15.

1.2.2 Maillage en points Collocaux

Nous utiliserons deux types de points collocaux : les points de Fekete et les points de Gauss-Lobatto dans le cadre de notre méthode pseudo-spectrale ou de collocation.

Les points de Fekete

Les points Fekete sont définis au moyen des fonctions de base de Dubiner [20]

$$\phi_{ij}(r,s) = \left(\frac{1-s}{2}\right)^i \times p_i^{0,0}\left(\frac{2r+s+1}{1-s}\right) \times p_j^{2i+1,0}(s)$$

où les $p_j^{\alpha,\beta}(s)$ sont les polynômes de Jacobi de degré j et d'ordre (α,β) [1]. Il est bien connu que l'ensemble des fonctions ϕ_{ij}, $0 \leq i,j \leq N$ et $i+j \leq N$ est une base orthogonale de $P_N(\widehat{T})$, l'espace du polynôme de degré inférieur ou égal à N sur un domaine triangulaire de référence, $\widehat{T}$. Dans tout ce qui suit, nous écrirons ϕ_k au lieu de ϕ_{ij}, $1 \leq k \leq (N+1)(N+2)/2$ pour toute bijection arbitraire $k \equiv k(i,j)$. Considérons maintenant la matrice de Vandermonde généralisé V dont les composants sont $V_{ij} = \phi_j(z_i)$ pour les points arbitraires $z_k \in \widehat{T}$, $k = 1,...,s$, où nous avons défini $s = (N+1)(N+2)/2$. Les points de Fekete sont les points $\widehat{z}_i$, $i = 1,...,s$ qui maximisent le déterminant de V :

$$\max_{\{z_i\}\in\widehat{T}} |V(z_1, z_2, ..., z_N)| \tag{1.1}$$

Les points de Gauss-Lobatto

Les points de Gauss-Lobatto correspondent aux racines des polynômes de Jacobie $p_j^{\alpha,\beta}(s)$. Cependant, dans le cas des domaines à produits tensoriels comme des droites ou des rectangles, les points de Gauss-Lobatto sont bien adaptés à l'approximation spectrale [15], [16]. Et dans le cas des domaines complexes, les points sont générés dans une élément rectangulaire standard premièrement et transférés par la suite

dans le domaine complexe en utilisant une transformation appropriée. Dans Figure 1.3, nous présentons des exemples de points de collocation générés sur un domaine complexe.

On peut remarquer que les points de collocation sont indépendants de la base choisie. En outre, comme il faut calculer numériquement l'inverse des matrices dans l'approximation des opérateurs de différentiations, il est important pour les matrices obtenues d'être bien conditionnées.

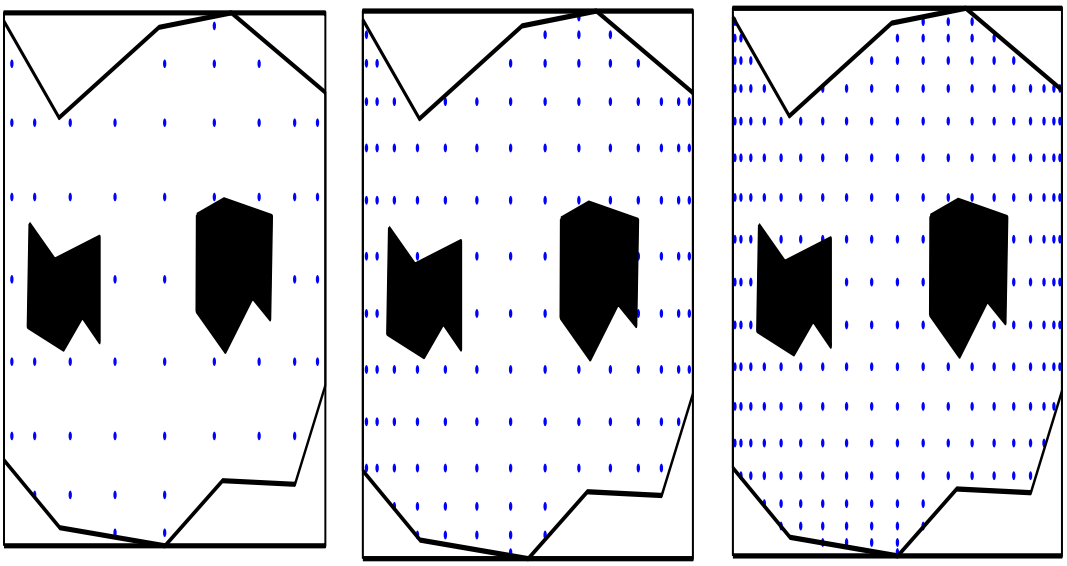


Figure 1.3 – *De la gauche vers la droite, les points de collocation construits respectivement pour* $NP = 10$*,* 15 *et* 20*.*

1.2.3 Maillage mixte

Le maillage mixte combine les éléments finis et les points de collocations. Le principal avantage de cette méthode, comparée aux travaux [46] et [47], est qu'elle recherche la meilleure valeur des paramètres de maillage NT (paramètre de triangularisation) et NP (Paramètre des Points de collocation) pour une meilleure approximation et conditionnement des matrices de différenciation. Pour cette raison,

la nouvelle approche reste efficace, même pour les petites valeurs de NT et NP.

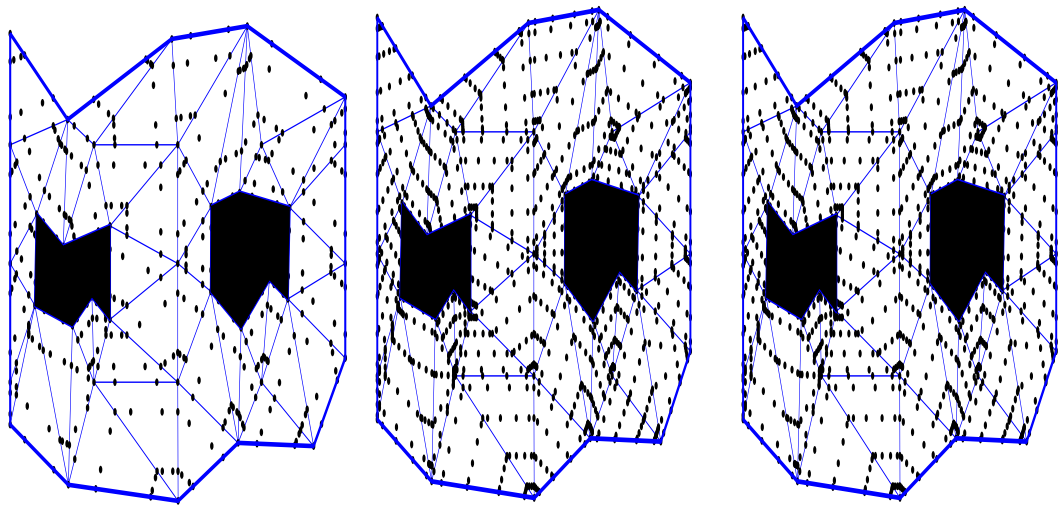


Figure 1.4 – *De gauche à droite, Les éléments finis Points de collocation pour respectivement les couples de paramètres ($NT = 5$, $NP = 3$), ($NT = 5$, $NP = 5$) et ($NT = 10$, $NP = 3$).*

1.3 Formulation de schémas numériques

Dans cette section nous faisons un bref rappel sur la construction des matrices de différentiations pseudo-spectrales définies sur des domaines non réguliers. A cet effet, nous emprunterons certaines notations de [46],[47] et [49] qui ont longuement étudié cet aspect.

1.3.1 Méthode de différentiation Pseudo-spectrale [38, 39, 40, 47]

L'approximation des opérateurs de différentiations par des matrices de différentiation pseudo-spectrale passe par un maillage du domaine d'étude en éléments finis et ensuite par des points de collocation générés sur lesdits éléments finis. Ces points

de collocation constituent des valeurs particulières pour des fonctions de base associées à chaque élément fini. Ainsi nous auront les points de Gauss-Lobatto qui sont issus des polynômes de Jocobie et des points de Feketes qui sont issus des fonctions de la base de Dubiner. Nous rappellerons dans un premier temps les opérateurs de différentiations ensuite la méthode de collocation moindres carrés et enfin la méthode itérative de Denor-cell.

Matrices de Différentiation

Dans cette sous-section, nous rappelons les techniques développées dans [47, 38, 39, 40] qui ont longuement travaillé sur les performances de ces matrices selon la nature du domaine et du type d'équation. Tout d'abord, l'ensemble du domaine est maillé par des éléments finis triangulaires. Deuxièmement, nous définissons les points de collocation dans un triangle standard en utilisant soit les points de Fekete [44] ou les points de Gauss-Lobatto [51] et enfin, par une transformation bijective, nous transférons les points de collocation du triangle de référence à un triangle arbitraire (voir la figure 1.5). Pour être plus compréhensif, considérons le triangle de référence défini par

$$\widehat{T} = \{(r,s), -1 \leq r,s \leq 1;\ r+s \leq 0\} \tag{1.2}$$

ainsi que la transformation bijective entre un triangle rectangle (standard), $\widehat{T}$ et un triangle quelconque T :

$$\begin{array}{rccl} h_1: & \widehat{T} & \to & T \\ & \begin{pmatrix} r \\ s \end{pmatrix} & \to & \begin{pmatrix} \dfrac{x_2-x_1}{2} & \dfrac{x_3-x_1}{2} \\ \dfrac{y_2-y_1}{2} & \dfrac{y_3-y_1}{2} \end{pmatrix} \times \begin{pmatrix} r \\ s \end{pmatrix} + \begin{pmatrix} \dfrac{x_2+x_3}{2} \\ \dfrac{y_2+y_3}{2} \end{pmatrix} \end{array} \tag{1.3}$$

Les (x_i, y_i), $i = 1, 2, 3$ sont les cordonnées des sommets du triangle T. Notons que toute fonction continue sur le triangle standard $\widehat{T}$ peut être approchée par $U(\widehat{\mathbf{x}}_i) \in$

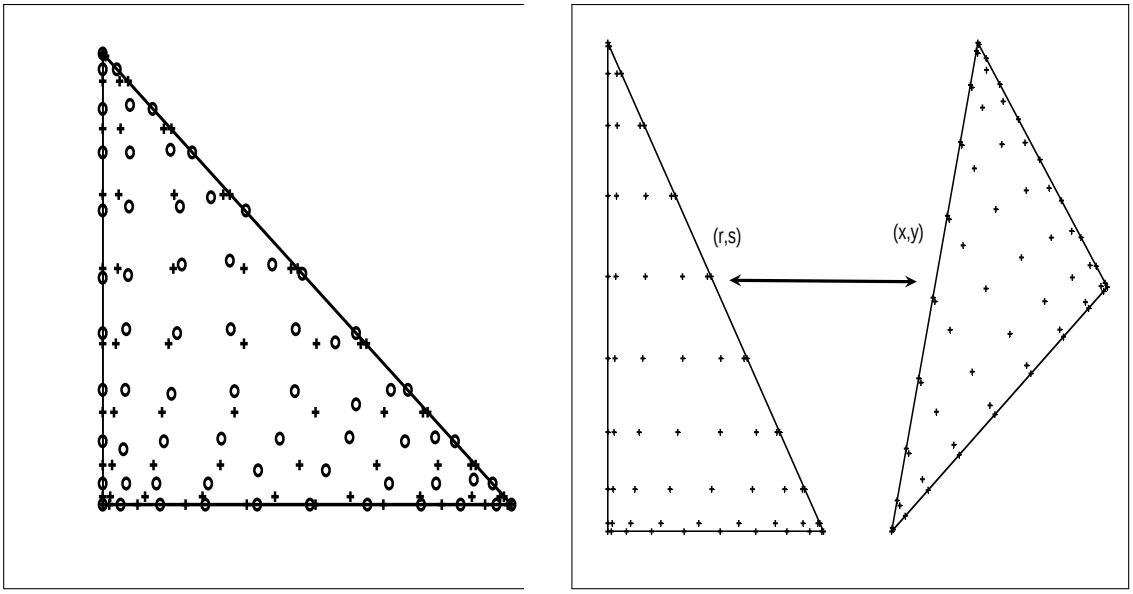


Figure 1.5 – *(A gauche) Les points de Gauss-Lobatto (+) et les points de Fekete (o) representés sur le triangle de référence* $\widehat{T}$ *pour* $N = 10.$; (A droite) *Transfert des points de Fekete du triangle de référence (gauche) au triangle arbitraire (droit) pour* $N = 10$.

$\mathcal{P}_N\ (\widehat{T})$ satisfaisant à

$$U(\widehat{\mathbf{x}}_i) = \sum_{k=1}^{d(N)} U_k \times \psi_k(\widehat{\mathbf{x}}_i),\ i = 1, .., d(N) \tag{1.4}$$

où $\{\widehat{\mathbf{x}}_k\}_{1\leq k\leq d(N)}$ est une suite de points collocaux définie sur $\widehat{T}$ et les U_k sont les coordonnées de la fonction approchée dans la base (voir [46] pour le calcul des U_k).

Appliquant les règles de dérivation dans les directions r et s on obtient :

$$\begin{aligned} \partial_r U_N(r,s) &= \sum_{k=1}^{d(N)} U_k \times \partial_r \psi_k(r,s), \\ \partial_s U_N(r,s) &= \sum_{k=1}^{d(N)} U_k \times \partial_s \psi_k(r,s) \end{aligned} \tag{1.5}$$

où $\mathcal{P}_N\ (\widehat{T})$ est l'espace de polynômes de degré inférieur ou égal à N, $d(N)$ est le nombre total de points collocaux générés dans $\widehat{T}$. Admettons par ailleurs les

notations matricielles suivantes :

$$\begin{array}{rcrcl} V & : & V_{ij} & = & \psi_j(\widehat{\mathbf{x}}_i) \quad \textit{(matrice de Vandermonde)} \\ V^r & : & V^r_{ij} & = & \partial_r\psi_j(\widehat{\mathbf{x}}_i) \\ V^s & : & V^s_{ij} & = & \partial_s\psi_j(\widehat{\mathbf{x}}_i) \\ V^{rs} & : & V^{rs}_{ij} & = & \partial_r\partial_s\psi_j(\widehat{\mathbf{x}}_i) \end{array} \tag{1.6}$$

et

$$\begin{array}{rcl} \widehat{U} & = & \Big(U_1, U_2, ..., U_{d(N)}\Big)^T \\ \widetilde{U}_r & = & \Big(U_r(\widehat{\mathbf{x}}_1), U_r(\widehat{\mathbf{x}}_1), ..., U_r(\widehat{\mathbf{x}}_{d(N)})\Big)^T \\ \widetilde{U}_s & = & \Big(U_s(\widehat{\mathbf{x}}_1), U_s(\widehat{\mathbf{x}}_1), ..., U_s(\widehat{\mathbf{x}}_{d(N)})\Big)^T \\ \widetilde{U} & = & \Big(U(\widehat{\mathbf{x}}_1), U(\widehat{\mathbf{x}}_1), ..., U(\widehat{\mathbf{x}}_{d(N)})\Big)^T \end{array} \tag{1.7}$$

Il faut noter que le choix des points de collocation est fait de telle sorte que la matrice de Vandermonde V soit inversible. Ce critère de selection a fait l'objet d'un algorithme développé dans [47]. Ainsi, Nous déduisons alors les matrices de différentiation d'ordre un et deux sur le triangle standard que sont :

$$\begin{array}{rcl} D^r & = & V^r \times V^{-1} \\ D^s & = & V^s \times V^{-1} \\ D^{rs} & = & V^{rs} \times V^{-1} \end{array} \tag{1.8}$$

de telle sorte que

$$\begin{array}{rcl} \widetilde{U_r} & = & D^r.\widetilde{U} \\ \widetilde{U_s} & = & D^s.\widetilde{U} \\ \widetilde{U_{rs}} & = & D^{rs}.\widetilde{U} \end{array} \tag{1.9}$$

On généralise cette définition en considérant la transformation ponctuelle (1.3) et on obtient pour un triangle quelconque les matrices de différentiations de premier

et de second ordre suivant :

$$\begin{aligned} D^x &= (c_{11} \times V^r + c_{21} \times V^s) \times V^{-1} \\ D^y &= (c_{12} \times V^r + c_{22} \times V^s) \times V^{-1} \\ D^{xx} &= (c_{11}^2 V^r + 2.c_{11}.c_{21} V^{rs} + c_{21}^2.V^s) .V^{-1} \\ D^{yy} &= (c_{12}^2 V^r + 2.c_{12}.c_{22} V^{rs} + c_{22}^2.V^s) .V^{-1} \\ D^{xy} &= (c_{12}.c_{11}) V^r + (c_{11}.c_{22} + c_{21}.c_{12}).V^{rs} + (c_{22}.c_{21}).V^s) .V^{-1} \end{aligned} \tag{1.10}$$

où $c_{i,j}$ sont des valeurs constantes.

Processus d'assemblage

L'avantage d'approximer les opérateurs de différentiation par une matrice de différentiation locale (appliquée à chaque élément triangulaire) est de former une matrice diagonale, de grande taille, par bloc contenant toutes les matrices de différentiations locales. Ainsi le système d'équations aux dérivées partielles peut se réécrire sous forme d'une équation différentielle ordinaire.

1.3.2 Méthode de collocation moindres carrés

Considérons le système d'équations (1.22) modélisant la dynamique de la pollution dans une eau de surface. En appliquant la méthode de discrétisation expliquée ci-dessus à cette équation, on obtient le système de différentiation ordinaire discrète suivant :

$$\begin{aligned} \frac{d\underline{C}(t)}{dt} &= \xi \mathbb{L} \times \underline{C}(t) - \eta \underline{C}(t) \\ &\quad -\mu \left|\underline{C}\right|^p (t) - \underline{f}(t;\lambda) \\ Z_1 \times \mathbb{D}\underline{C}(t) &= Z_1 \times \underline{\Phi}_1(t) \\ Z_2 \times \mathbb{D}\underline{C}(t) &= Z_2 \times \underline{\Phi}_2(t) \\ \underline{C}(0) &= \underline{C^0}(x) \end{aligned} \tag{1.11}$$

où $t \in]0, T]$, $\mathbb{D}$ la matrice de différentiation du bord et :

$$\underline{C}(t) = \Big(C(t, \mathbf{x}_1), ..., C(t, \mathbf{x}_{d(N)})\Big)' ; \tag{1.12}$$

$$\underline{f}(t; \lambda) = \Big(f(t, \mathbf{x}_1; \lambda), ..., f(t, \mathbf{x}_{d(N)}; \lambda)\Big)' ; \tag{1.13}$$

$$\underline{g}(t; \tau) = \Big(g(t, \mathbf{x}_1; \tau), ..., g(t, \mathbf{x}_{d(N)}; \tau)\Big)' ; \tag{1.14}$$

$$\underline{\Phi_1}(t; \tau) = \Big(\Phi_1(t, \mathbf{x}_1; \tau), ..., \Phi_1(t, \mathbf{x}_{d(N)}; \tau)\Big)' ; \tag{1.15}$$

$$\underline{\Phi_2}(t; \tau) = \Big(\Phi_2(t, \mathbf{x}_1; \tau), ..., \Phi_2(t, \mathbf{x}_{d(N)}; \tau)\Big)' ; \tag{1.16}$$

les $\mathbf{x}_k$, $k = 1, ..., d(N)$ sont les points de collocation. Z_i, $i = 1; 2$ désigne l'opérateur discret qui identifie les points du bord du domaine Γ_i. L'erreur commise dans l'approximation du système continue par le système discret (1.11) est estimée par l'opérateur suivant :

$$\begin{aligned} \mathscr{L}(\underline{v}, \underline{C}) \;\; = \;\; & \left\| \frac{d\underline{C}(t)}{dt} - \xi \mathbb{L} \times \underline{C}(t) + \eta \underline{C}(t) \right. \\ & \left. - \mu \left| \underline{C} \right|^p (t) - \underline{f}(t; \lambda) \right\|^2 . \end{aligned} \tag{1.17}$$

Notons le vecteur des paramètres inconnus par

$$\underline{v} = (\xi, \eta, \mu, p, \lambda, \tau). \tag{1.18}$$

Pour évaluer $\underline{v}$ nous avons besoin de données observées de C solution de l'équation (1.22) sur tout l'étendu du domaine d'étude Ω. Dans la pratique, c'est une tâche difficile d'avoir de telles données.

Dans ce paragraphe, nous sommes concernés par le problème inverse suivant :

Ayant des données collectées sur la solution de notre équation, nous voulons estimer $\underline{v}$ et la solution numérique $\underline{C}(t)$ dans l'intégralité du domaine. Pour résoudre ce problème, nous nous sommes inspirés des travaux sur la méthode des sentinelles dévéloppés par J.L. Lions [41] et A. Traore [48]. A cet effet, notons par C_{obs} lesdites données et $\Omega_{obs} \subset \Omega$ le domaine d'observation, un sous domaine de Ω (Figure 1.7) où les mesures expérimentales ont été prélevées. Nous désignons par Z_{obs} et Z_i $(i = 1; 2)$ les matrices associées aux points de collocations du domaine d'observation respectivement du bord du domaine Γ_i tel que :

$$\begin{aligned} Z_{obs} \times \underline{C} &= C|_{x\in\Omega_{obs}} \\ Z_i \times \underline{C} &= C|_{x\in\Gamma_i}. \end{aligned} \tag{1.19}$$

Nous cherchons à évaluer $\underline{C}(t, \underline{v})$ solution de (1.11) et le vecteur de paramètres inconnus (1.18). Nous résumons la formulation de la méthode de collocation-moindres carrées de (1.11) comme le suivant :

$$\begin{aligned} &\text{identifier } \underline{v}^*, \underline{C}^*, \quad \text{solutions de} \\ &J_{\widehat{\beta}}(\underline{v}^*, \underline{C}^*) = \min_{(\underline{v},\underline{C})\in\mathbb{R}^6\times\mathbb{R}^{d(N)}} J_{\widehat{\beta}}(\underline{v}, \underline{C}) \end{aligned} \tag{1.20}$$

où

$$\begin{aligned} J_{\widehat{\beta}}(\underline{v}, \underline{C}) &= \sum_{k=1}^{[T/\Delta t]} \|Z_{obs} \times \underline{C}(k\Delta t) - \underline{C}_{obs}(k\Delta t)\|^2 \\ &+ \sum_{k=1}^{[T/\Delta t]} \left\|Z_1\mathbb{D} \times \underline{C}(k\Delta t) - Z_1 \times \underline{\Phi_1}(k\Delta t)\right\|^2 + \sum_{k=1}^{[T/\Delta t]} \left\|Z_2\mathbb{D} \times \underline{C}(k\Delta t) - Z_2 \times \underline{\Phi_2}(k\Delta t)\right\|^2 \\ &+ \widehat{\beta} \times \|\underline{v} - \widehat{\underline{v}}\|^2 + \mathscr{L}(\underline{v}, \underline{C}) \end{aligned} \tag{1.21}$$

et $\underline{C}_{obs}(t)$ est un vecteur de mesure obtenu au point $\mathbf{x} \in \Omega_{obs}$ et au temps t. Une contrainte positive est établie sur $\underline{C}$. $\widehat{\beta}$ est un paramètre de régularisation de Tikhonov et $\widehat{\underline{v}}$ est une information à priori. Dans cet exemple nous cherchons des valeurs de $\underline{v} \in]0, 2]$.

1.3.3 Méthode itérative de Denor-cell

Le schéma numerique de Denor-cell a été développé par Griebel et al [24] pour l'approximation des opérateurs de différentiation pour l'étude d'écoulement des fluides dans un canal modélisé par des équations de Navier-Stokes. Pour illustrer un tel schéma :

Considérons le domaine régulier $\Omega = [a, b] \times [c, d] \in \mathbb{R}$ sur lequel nous introduisons une grille de points (x_i, y_j) telle que $x_{i+1} = x_i + \delta x,\ i = 0, ..., i_{max} - 1$ et $y_{j+1} = y_j + \delta y, j = 0, ..., j_{max} - 1$ où $\delta x = (b - a)/i_{max}$ et $\delta y = (d - c)/j_{max}$ sont les pas de discrétisation suivant x et y respectivement. Le domaine Ω est ainsi discrétisé en $(i_{max} \times j_{max})$ cellules (Figure 1.6). Les cellules sont numérotées dans l'ordre lexicographique (de la gauche vers la droite et du bas vers le haut).

Enfin, pour tous $i = 0, ..., i_{max} - 1,\ j = 0, ..., j_{max} - 1$ nous introduisons le point $X_{i,j} = (x_{i+1/2}, y_{i+1/2})$ de Ω comme étant le centre de la cellule (i, j) (Figure 1.6). Considérons une fonction $h = h(x, y)$ assez régulière définie sur Ω. Étant données des valeurs de h aux points $X_{i,j}$.

Dans la suite, On définit les matrices de différentiations spatiales du premier et second ordre. $H_{i,j}$ désigne la valeur de h au point $X_{i,j}$.

Matrices de différentiations : différences finies centrales

L'approximation des opérateurs $\frac{\partial h}{\partial x}$ et $\frac{\partial h}{\partial y}$ à l'ordre $o(\delta x)$ donne :

$$\frac{\partial h}{\partial x}(X_{i,j}) \approx \frac{1}{\delta x}(H_{i+1,j} - H_{i-1,j})$$

et

$$\frac{\partial h}{\partial y}(X_{i,j}) \approx \frac{1}{\delta y}(H_{i+1,j} - H_{i-1,j}).$$

Matrice de différentiations : différences finies progressives

Nous définissons, comme précédemment, les opérateurs de premiers ordre comme suit :

$$\frac{\partial h}{\partial x}(X_{i,j}) \approx \frac{1}{\delta x}(H_{i+1,j} - H_{i,j})$$

et

$$\frac{\partial h}{\partial y}(X_{i,j}) \approx \frac{1}{\delta y}(H_{i+1,j} - H_{i,j}).$$

Matrice de différentiation : Schema de Denor-cell

Le schema de Denor-cell a été développé pour l'approximation des opérateurs de second ordre. Il s'apparente à un couplage des deux précédentes méthodes. Soit $K(h)$ une application régulière et sa valeur au point $X_{i,j}$ est donnée par $K_{i,j}$.

$$\frac{\partial}{\partial x}\left(K(h)\frac{\partial h}{\partial x}\right)(X_{i,j}) \approx \frac{1}{\delta x}\left(K_{i+\frac{1}{2},j}\frac{H_{i+1,j} - H_{i,j}}{\delta x} - K_{i-\frac{1}{2},j}\frac{H_{i,j} - H_{i-1,j}}{\delta x}\right)$$

et

$$\frac{\partial}{\partial y}\left(K(h)\frac{\partial h}{\partial y}\right)(X_{i,j}) \approx \frac{1}{\delta y}\left(K_{i,j+\frac{1}{2}}\frac{H_{i,j+1} - H_{i,j}}{\delta x} - K_{i,j-\frac{1}{2}}\frac{H_{i,j} - H_{i,j-1}}{\delta x}\right).$$

1.4 Expérimentation numérique

1.4.1 Étude de cas 1 : Pollutions dans une eau de surface.[46]

Mise en équation

En prenant en compte les propriétés des fluides, nous décrivons une propagation de la pollution dans une eau de surface, $C(t, \mathbf{x})$, comme suit :

Diffusion : Un terme de diffusion $kdiv(a\nabla C(t, \mathbf{x}))$. Ici k est la constante de diffusion et a est la constante de transmissibilité dans le milieu.

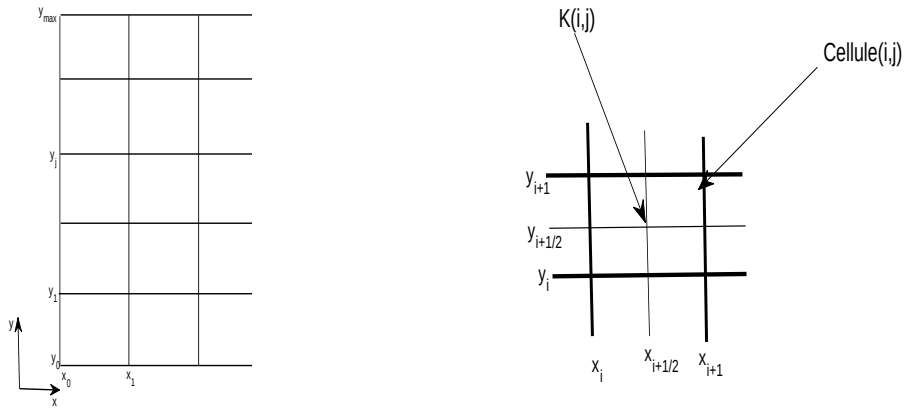


Figure 1.6 – *(A gauche) Maillage d'un domaine tensoriel* Ω *en* $i_{max} \times j_{max}$ *cellules. (A droite) représentation d'une cellule* (i, j) *dans la grille.*

Transport : Un terme de transport ou convection $\overrightarrow{u}\nabla C(t, \mathbf{x})$, où $\overrightarrow{u}$ désigne le champ de vitesse du fluide. Ce terme prend donc en compte l'effet de l'écoulement du fluide et dans l'étude d'un lac nous considérerons $\overrightarrow{u} = \overrightarrow{0}$ du fait de l'absence d'écoulement de l'eau. Ce qui n'est pas le cas d'une rivière.

Réaction : Un terme de réaction R qui traduit les interactions chimiques et biochimiques dans le liquide. Ces interactions se déroulent entre les micro-organismes d'une part et les composés chimiques d'autre part. Nous nous sommes intéressés à une réaction de type explosive donnée sous la forme $-\lambda C(t, \mathbf{x}) + \mu |C|^p (t, \mathbf{x})$. Ce choix a été motivé par les réactions d'oxydation qui se produisent dans un fluide pollué(voir [48])

Source de pollution : La modélisation de la source de pollution s'est inspirée des travaux de J.P. Kernevez dans le cas d'un fleuve, [30]. Ainsi, la source de pollution est décrite par une fonction $f(t, \mathbf{x})$. Elle donne naissance aux substances polluantes déversées dans le fluide. A ce niveau, deux considérations sur les sources de pollution méritent d'être faites pour une meilleure prise en compte des espèces polluantes :

Il s'agit des termes sources distribués que nous noterons par $\xi(t, \mathbf{x})$ et les

termes sources ponctuels au point i de cordonnée $\mathbf{x}_i$ que nous désignerons par $\lambda_j\widehat{\xi}_i(t) \times \delta(\mathbf{x} - \mathbf{x}_i)$. La formulation générale de la source est donnée par

$$f(t, \mathbf{x}) = \xi(t, \mathbf{x}) + \sum_i \lambda_j\widehat{\xi}_i(t) \times \delta(\mathbf{x} - \mathbf{x}_i)$$

où $\delta(\mathbf{x} - \mathbf{x}_i)$ représente la fonction de Dirac associée à $\mathbf{x}_i$.

Conditions aux Bords : Le bord du domaine, Γ, considéré dans ce problème peut être scindé en fonction de la nature des échanges qui ont lieu au contact du fluide(l'eau) et les parois du domaine(la nappe phréatique). Ici, nous considérerons deux types Γ_1 et Γ_2. Soit $\Gamma = \Gamma_1 \cup \Gamma_2$:

(i)- En Γ_1, on note une absence d'échange de pollution entre les deux milieux et la concentration de la pollution reste équivalente à une fonction $\Phi_1(x, t),\ x \in \Gamma_1$;

(ii)- En Γ_2, le flux d'échange de pollution à travers la nappe phréatique est matérialisé par la fonction $\Phi_2(x, t),\ x \in \Gamma_2$.

En somme les conditions aux bords se résument en :

$$\begin{aligned} C(x, t) &= \Phi_1(t, \mathbf{x}), \quad x \in \Gamma_1; \\ \frac{\partial C(x, t)}{\partial n} &= \Phi_2(t, \mathbf{x}), \quad x \in \Gamma_2 \end{aligned}$$

et la

Condition initiale :

$$C(0, \mathbf{x}) = g(\mathbf{x}; \tau).$$

Nous résumons le transport d'une unité de concentration dans une eau de surface

par le système parabolique suivant :

$$\begin{aligned}
\frac{\partial C(t,\mathbf{x})}{\partial t} &= \xi\Delta C(t,\mathbf{x}) + \eta C(t,\mathbf{x})) - \mu\left|C\right|^{p}(t,\mathbf{x})) \\
&\quad + f(t,\mathbf{x};\lambda), \qquad (t,\mathbf{x}) \in]0,T]\times\Omega, \\
C(t,\mathbf{x}) &= \Phi_1(t,\mathbf{x}), \qquad (t,\mathbf{x}) \in]0,T]\times\Gamma_1, \qquad \textit{Absence de flux} \\
\frac{\partial C(t,\mathbf{x})}{\partial \mathbf{n}} &= \Phi_2(t,\mathbf{x}), \qquad (t,\mathbf{x}) \in]0,T]\times\Gamma_2, \qquad \textit{Présence de flux} \\
C(0,\mathbf{x}) &= g(\mathbf{x};\tau), \qquad \mathbf{x}\in\Omega
\end{aligned} \tag{1.22}$$

où

le domaine : $\Omega \subset \mathbb{R}^2$;

la durée de l'expérience : $]0,T]$, $T>0$, est la durée de l'expérience;

les paramètres inconnus : Les réels ξ, η, μ, p, λ et τ sont des réels strictement positifs dont la valeur de chacun est inconnue dans le système d'équations.

Nous sommes concernés par un problème d'identification de paramètres qui consiste à déterminer les paramètres inconnus ξ, η, μ, p, λ et τ dans le système (1.22). Pour les aspects théoriques concernant l'existence et l'unicité d'une solution d'un tel système (1.22), le lecteur pourra voir les documents suivants : [21], [32] et [41].

Les données du problèmes

Pour le cas d'un Lac, nous avons considéré que $\Gamma = \Gamma_1$ et que nous sommes dans le cas d'une absence de flux et $\vec{u} = \vec{0}$. De plus, l'expression de la fonction source est donnée par :

$$f(t,\mathbf{x}) = 10\lambda sin(\pi t + 1)\delta_{\mathbf{x}_c}(\mathbf{x})$$

où $\delta_{\mathbf{x}_c}(\mathbf{x})$ est une fonction de Dirac définit au point $\mathbf{x}_c$, où $\mathbf{x}_c = (0, 0)$, tel que

$$\delta_{\mathbf{x}_c}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{x}_c \\ 0 & \text{ailleurs.} \end{cases}$$

La fonction initiale est exprimée par

$$g(x) = \tau. \left(x^2 + y^2 - .06\right) \delta_{\mathbf{x}_c}(\mathbf{x}).$$

Collecte des données : Pour tester notre algorithme, nous avons construit des données en laboratoire, $\underline{C}_{obs}(t)$ comme suit :

Figure 1.9 : Premièrement, nous considérons des courbes d'évolution de la concentration en pollution prélevée en cinq points, O_i, $i = 1, ..., 5$ (points d'observation, Figure 1.7) situés dans Ω_{obs},

Figure 1.8 : ensuite, à partir de 24 points pris sur l'axe des abscisses représentant l'intervalle de temps (par jour), nous constituons les données de concentration prélevées dans chaque puit O_i par projection sur les courbes respectives. Finalement, nous recouvrons le maximum de données prélevées sur tous les points de collocation inclus dans Ω_{Obs}.

Grâce au logiciel de programmation scientifique, MATLAB 7.01, qui contient des algorithmes intégrés. Ainsi, pour la résolution de (1.20) nous avons utilisé le solver " *fmincon*" et pour (1.17) nous avons utilisé le "solver" "*ode45*".

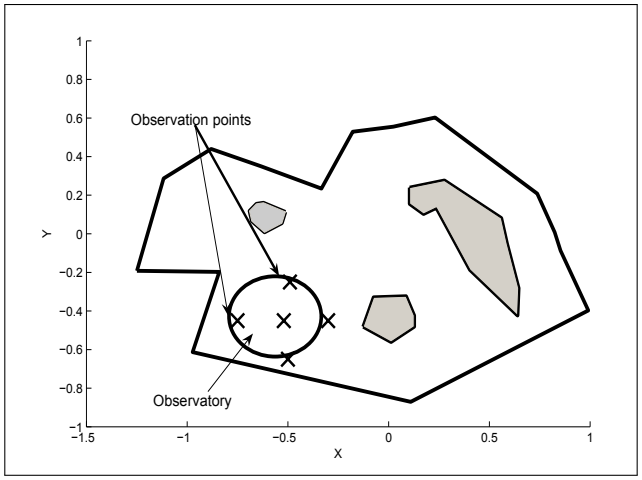


Figure 1.7 – *Une vue du domaine d'observation Ω_{obs} et les points, O_i utilisés pour les mesures : Cas d'un Lac.*

Résultats : En utilisant les données dans la figure 1.8, pour $0 \leq t \leq 24$ et après simulation de (1.20), les valeurs optimales de $\underline{v}$ obtenus sont :

$\xi =$	58.10^{-4}
$\eta =$	10^{-6}
$\mu =$	6
$p =$	0.3373
$\lambda =$	0.4999
$\tau =$	1.4680

Les valeurs approximatives des paramètres constituent des résultats importants dans le processus de résolution numérique. Ils identifient clairement l'équation parabolique (1.22) :

L'expression de la fonction source f est déterminée :

$$f(t,\mathbf{x}) = 10\lambda sin(\pi t + 1)\delta_{\mathbf{x}_c}(\mathbf{x}) \quad avec \quad \lambda = 0.4999 \tag{1.23}$$

	mésures			
	O1	O2	O3	O4
t_1	0	0	0	0
t_2	0.0341	1.6315	2.4895	2.4303
t_3	0.1227	0.2481	1.5021	0.4914
t_4	0.2471	0.5347	1.0590	1.6734
t_5	0.4863	0.6602	0.7808	1.2885
t_6	0.8996	2.0906	2.6617	4.3541
t_7	1.9562	3.1453	2.7454	3.3367
t_8	4.7598	4.7221	5.4121	4.7685
t_9	8.0019	8.3292	9.1448	9.5251
t_{10}	10.1644	10.3390	11.0510	10.0418
t_{11}	10.2478	10.0611	11.3513	11.3318
t_{12}	9.9489	10.6747	11.3433	11.2951
t_{13}	11.0906	10.5023	11.6931	11.6932
t_{14}	11.6184	13.8016	12.5992	12.2813
t_{15}	11.0771	10.9407	10.9209	12.0159
t_{16}	12.0918	12.2057	12.0490	10.1750
t_{17}	14.0955	15.1623	13.5582	13.9864
t_{18}	14.1742	14.2334	14.4907	15.3864
t_{19}	15.1680	15.0724	14.0159	14.7469
t_{20}	15.3715	14.5391	15.9543	16.5321
t_{21}	16.5278	16.8222	16.0171	16.0574
t_{22}	16.0663	14.7302	15.2589	15.9360
t_{23}	16.2276	16.9419	17.1612	17.7301
t_{24}	14.0069	15.6304	14.7085	14.4529

Figure 1.8 – *Les données obtenues à partir de quatre points d'observation associés à la figure 1.9.*

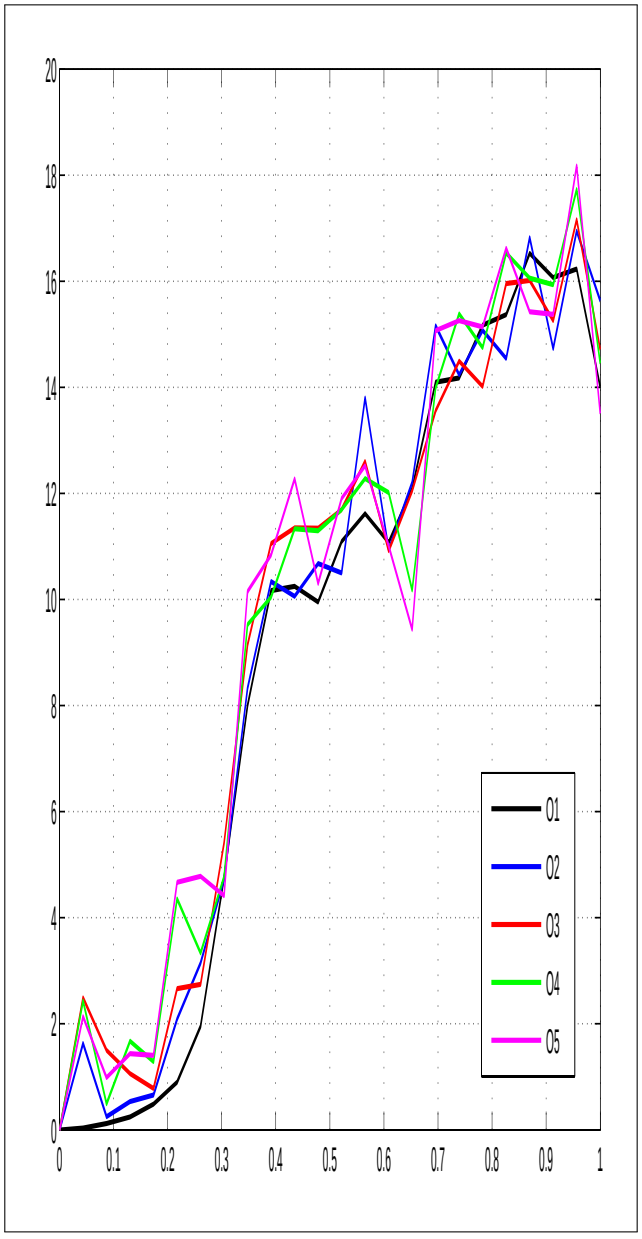


Figure 1.9 – *Les courbes des mesures sur points d'observation,* O_1, O_2, O_3, O_4 et O_5.

la fonction initiale est déterminée

$$g(x) = \tau.\,(x^2 + y^2 - .06)\,\delta_{\mathbf{x}_c}(\mathbf{x}) \quad avec \quad \tau = 1.4680$$

et le système d'équations qui gouverne notre modèle de pollution est établi :

$$\begin{aligned} \frac{\partial C(t,\mathbf{x})}{\partial t} &= \xi\Delta C(t,\mathbf{x}) + \eta C(t,\mathbf{x})) - \mu\,|C|^p\,(t,\mathbf{x})) \\ &\quad + f(t,\mathbf{x};\lambda), \qquad (t,\mathbf{x}) \in]0,T] \times \Omega, \\ C(t,\mathbf{x}) &= 0, \qquad (t,\mathbf{x}) \in]0,T] \times \Gamma, \end{aligned} \tag{1.24}$$

avec

$$\xi = 58.10^{-4}; \quad \eta = 10^{-6}; \quad \mu = 6 \quad et \quad p = 0.3373. \tag{1.25}$$

On peut approcher la solution de (1.24) par tout schéma numérique convenable d'équation différentielle partielle. La solution approximative est notée par $\underline{C}^*$.
Lorsque le temps évolue de 0 à 20, nous construisons en deux dimensions d'espace (Figures 1.11 et 1.12) et en dimension trois d'espace (Figure 1.10) la représentation graphique de la solution approchée $\underline{C}^*$. Ainsi nous pouvons apprécier l'évolution des solutions expérimentales sur les quatre images prises aux temps : $t = 2$, $t = 4$, $t = 6$ et $t = 20$. Nous faisons les observations suivantes :

Figure 1.10 : Premièrement, Nous remarquons la croissance de la quantité de concentration lorsque le temps de l'expérience croit. Ensuite, la forme de la solution contient des trous situés aux mêmes cordonnées que les obstacles introduits dans le domaine. Ceci signifie, qu'il n'y a pas de pollution dans ces lieux et il confirme notre hypothèse. La croissance de $\underline{C}^*$ est cohérent avec l'évolution des données (Figure 1.9). Enfin, à tout moment, nous remarquons dans les graphiques un pic de la concentration, c'est la valeur la plus grande de la concentration, elle indique la position de la source comme cela a été prédit dans Figure 1.23.

Figures 1.11 et 1.12 : nous décidons de représenter en dimension $2D-$ pour mettre en évidence la valeur de $\underline{C}^*$ autour des bords du domaine. Lorsque le temps croit, nous remarquons que la valeur de la concentration de pollution augmente. Nous notons également que la valeur de la concentration de pollution autour des bords du domaine(le périmètre de l'étendu d'eau et des obstacles à l'intérieur) augmente. Ceci est clairement visible dans la figure 1.11. Cette augmentation de la pollution se matérialise dans la figure 1.12, ainsi la couleur du domaine change du bleu (à $t = 2$) au vert (à $t = 20$). La position de la source est visible à travers la plus grande valeur, la couleur rouge et la surface des obstacles est maintenue à la couleur bleue, correspondante à la valeur nulle. Et nous remarquons une fois encore, la position de la source de pollution à travers chaque graphique.

Figure 1.13 : Nous sommes intéressés par la direction et les sens de déplacement des polluants, ainsi nous avons représenté le gradient de $\underline{C}^*$ en dimension 2. Les flèches indiquent les sens de déplacements des polluants dans le liquide. Lorsque le temps croit, les flèches se déplacent de la source vers les bords du domaine. Ceci justifie l'accumulation de polluants aux limites du domaine dans Figure 1.11.

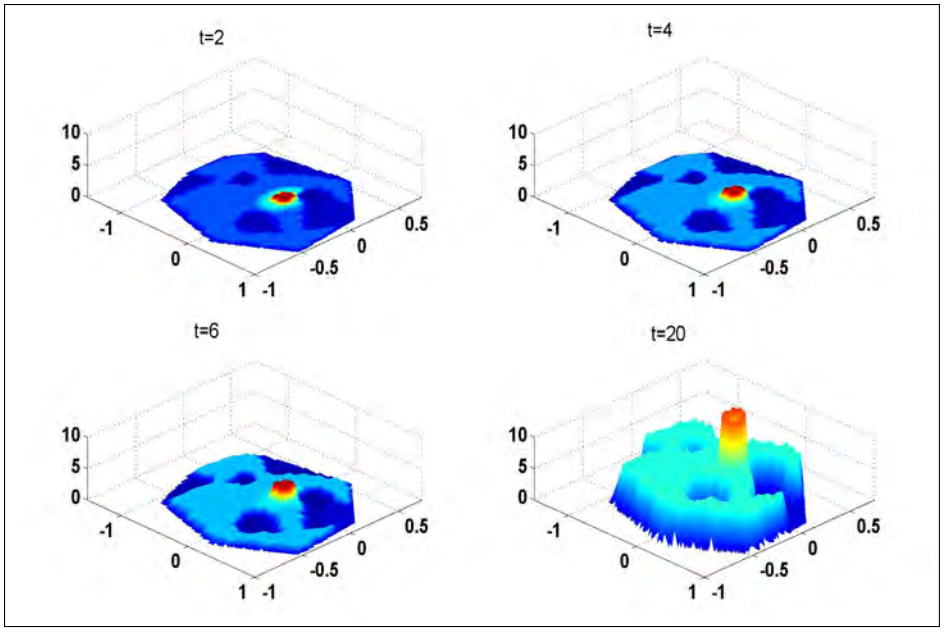


Figure 1.10 – *Schéma de la pollution en dimension 3D aux temps* $t = 2$, $t = 4$, $t = 6$ *et* $t = 20$.

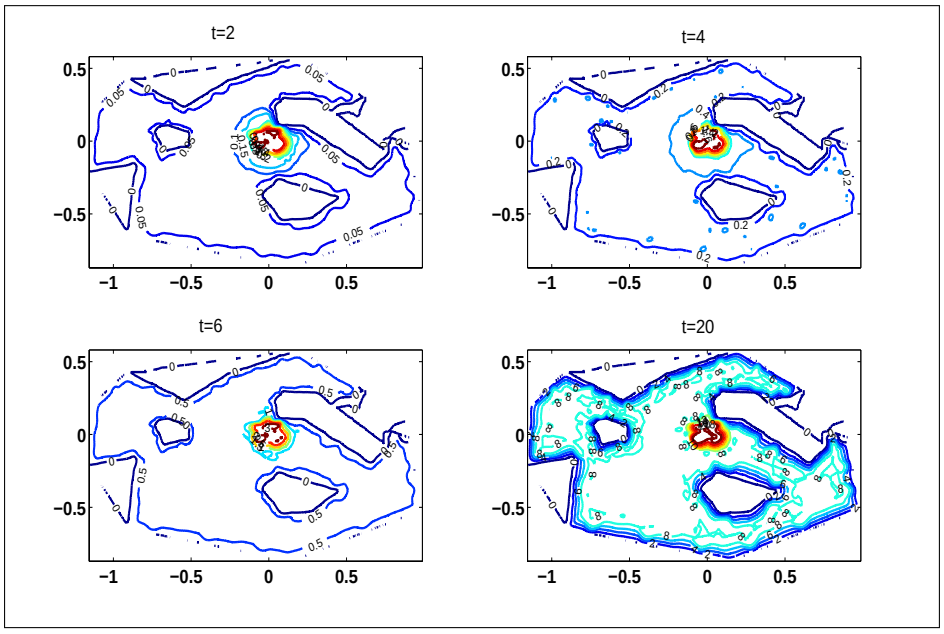


Figure 1.11 – *Présentation de la pollution aux temps* $t = 2$, $t = 4$, $t = 6$ *et* $t = 20$.

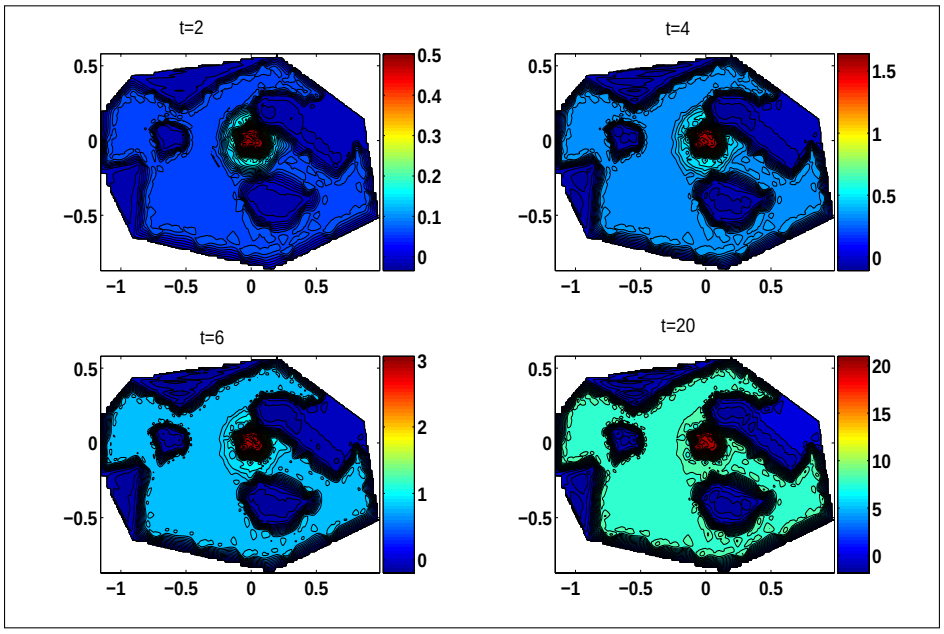


Figure 1.12 – *Présentation de la pollution aux temps* $t = 2$, $t = 4$, $t = 6$ *et* $t = 20$.

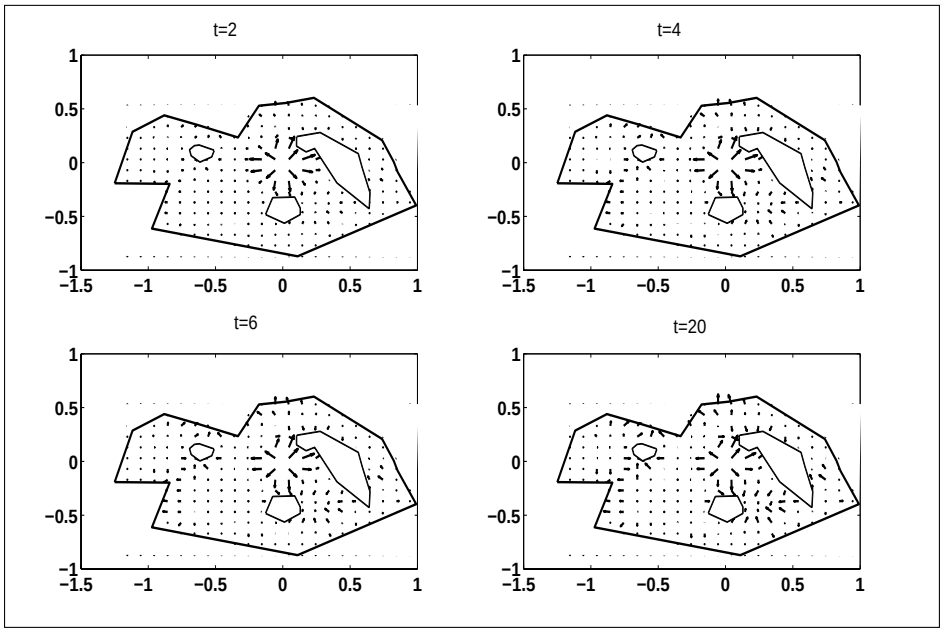


Figure 1.13 – *Présentation du sens de déplacement de la pollution aux temps* $t = 2$, $t = 4$, $t = 6$ *et* $t = 20$.

1.4.2 Étude de cas 2 : Assèchement d'un lac. [31]

Cette étude a fait l'objet d'une publication en 2013 [31]. Elle s'inspire du problème de l'assèchement du Lac Tchad pour développer un modèle mathématique et proposer une solution numérique en se basant sur des images satellitaires. En effet, considérant les images de Figure 1.14,

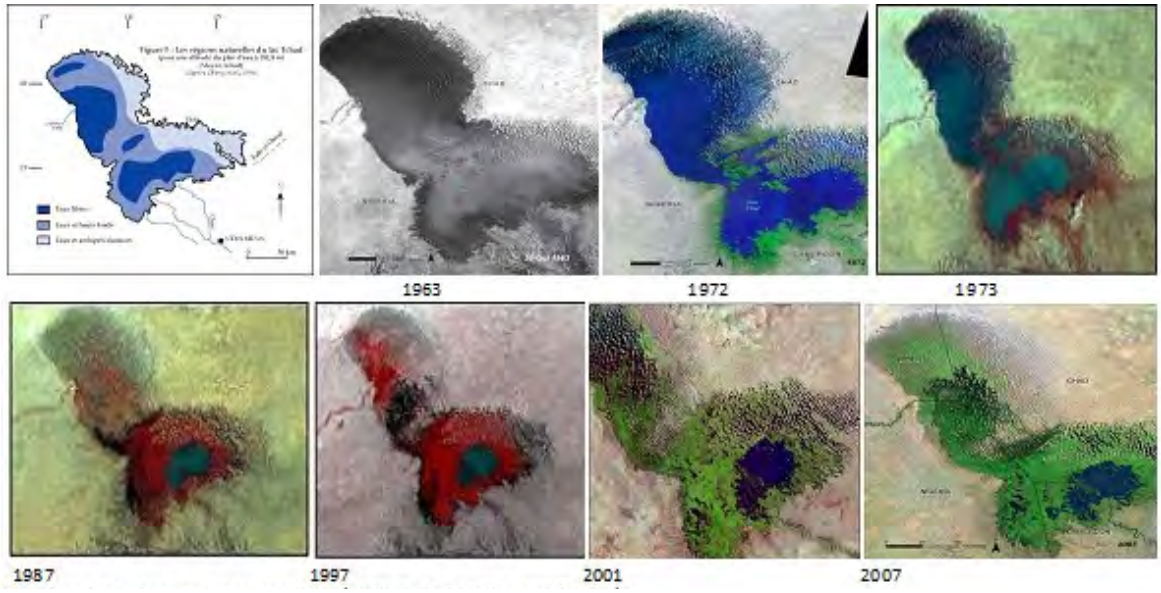


Figure 1.14 – *Evolution de l'assèchement du Lac Tchad de* 1963 *à* 2007. *Source : Nasa Goddard Space Flight Center.*

Mise en équation

Les lacs sont des systèmes dynamiques complexes en raison des paramètres hydrologiques (pluie, évaporation, flux entrant et sortant) et des caractéristiques physiques (le bassin, le contour). Ces paramètres peuvent changer d'un lac à l'autre. Dans l'article [31], nous avons concentré notre modèle sur les caractéristiques physiques du Lac Tchad pouvant être adaptées à d'autres lacs. Le modèle suit les lois physiques et les relations constitutives des principes de modélisation développés dans [26].

Supposons Ω un domaine borné de $\mathbb{R}^2$ qui représente la surface du lac dont le bord est noté par $\partial\Omega$ et supposé être assez régulier. De plus, soit $h(\mathbf{x}, t)$ la fonction qui décrit la hauteur de l'eau à la position $\mathbf{x} \in \Omega$ et au temps t. Soit V une unité de volume d'eau autour de $\mathbf{x}$, nous définissons l'équation suivante au point x et au temps t par

$$\frac{d}{dt}h(\mathbf{x}, t) + E(\mathbf{x}, t) = P(\mathbf{x}, t) + \frac{V(\mathbf{x}, t)}{A(h)} + \epsilon(t) \tag{1.26}$$

où h est le niveau du lac, A est la surface dépendante de la profondeur du lac, P est le taux de précipitations sur le lac, E est le taux d'évaporation du lac, t est le temps et V est le volume d'eau échangé (les entrées, les sorties d'eau dans le Lac). Le terme final ϵ représente les incertitudes dans le bilan hydrologique résultant d'erreurs dans les données et des instruments de mesures.

Aussi, nous supposons que les paramètres hydrologiques ne sont pas distribués uniformément, de sorte que les paramètres de modélisation E, P dépendent également de la variable de position.

Soit $\mathbf{q}(\mathbf{x}, t)$ le flux d'eau échangé tel que la quantité d'eau dans une unité de surface normale $\mathbf{n}$ est :

$$V(\mathbf{x}, t) = \mathbf{q}(\mathbf{x}, t) \cdot \mathbf{n} \tag{1.27}$$

par unité de surface ($A(h) = 1$). Ainsi, nous déduisons la loi de conservation de la masse pour V comme suit :

$$\int_V \frac{d}{dt}h(\mathbf{x}, t)d\mathbf{x} + \int_V E(\mathbf{x}, t)d\mathbf{x} = \int_V P(\mathbf{x}, t)d\mathbf{x} + \int_{\partial V} \mathbf{q}(\mathbf{x}, t) \cdot \mathbf{n}\, d\mathbf{x} + \epsilon(t) \tag{1.28}$$

où ∂V est le bord de V.

Dans la suite, nous utilisons le théorème de Green sur les intégrales de surfaces et comme v est arbitraire l'équation 1.28 nous donne

$$\frac{d}{dt}h + E = P + \nabla \cdot \mathbf{q} + \epsilon(t) \tag{1.29}$$

A ce stade, nous devons exprimer les paramètres E et $\mathbf{q}$ en fonction de h. Nous faisons les suppositions suivantes

$$E = \alpha h^r, 0 \leq r \leq 1, \tag{1.30}$$

où α et r sont des paramètres réels inconnus et peuvent être identifiés. Le flux $\mathbf{q}$ est proportionnel au volume par rapport au niveau. L'expression générale du flux est donnée par l'équation de transport et, à ce stade, l'hypothèse sur le coefficient de transport nous échappe, car $\mathbf{q}$ est l'agrégation des flux de surface et du sol :

$$\mathbf{q} = k(\mathbf{x}, t)\nabla h \tag{1.31}$$

où $k(\mathbf{x}, t) = ah(\mathbf{x}, t)(a > 0)$ est le coefficient de transport de l'eau, ∇h est la variation du niveau d'eau dans les directions x et y. Ce terme permet de prendre en compte le processus d'ensablement du lac ou tout élément géologique contribuant à réduire le niveau d'eau au fond du bassin. Enfin, nous obtenons le système d'équations suivant pour décrire le niveau du lac :

$$\frac{\partial h}{\partial t} - \nabla \cdot \Big(k(h)\ \nabla h\Big) + \alpha\ h^r = P \quad \text{dans} \quad \Omega \times]0,\ \infty) \tag{1.32a}$$

$$h = 0 \quad \text{dans} \quad \partial\Omega \times [0,\ \infty) \tag{1.32b}$$

$$h(\mathbf{x}, 0) = h_0(\mathbf{x}) > 0,\ \forall \mathbf{x} \text{ dans } \overline{\Omega} = \Omega \cup \partial\Omega. \tag{1.32c}$$

Le système (1.32) est un cas spécial du modèle introduit par J.I Diaz dans [18].

Pour des raisons d'approximation numérique et éviter que le problème (1.32) soit mal posé, nous admettrons les hypothèses suivantes :

— $h \in C^1(0,\infty; L^2(\Omega))$;

— k, α et r sont des constantes positives, les paramètres du modèle ;

— $h_0 \in L^2(\Omega)$, et $h_0 = 0$ dans $\partial\Omega$;

— $P \in C^1([0,\infty]),\ P \geq 0$;

Sous ces hypothèses de regularité, le système $(8a) - (8c)$ est bien posé [18].

Discrétisation du modèle

L'équation $(1.32a)$ peut être écrite comme suit :

$$\frac{\partial h}{\partial t} - \frac{\partial}{\partial x}\Big(k(h)\frac{\partial h}{\partial x}\Big) - \frac{\partial}{\partial y}\Big(k(h)\frac{\partial h}{\partial y}\Big) + \alpha\, h^r = P \tag{1.33}$$

En appliquant le schéma de Donor-cell en espace pour l'équation (1.32a) cela conduit à la semi-discrétisation suivante :

$$\begin{aligned}\left(\frac{\partial h}{\partial t}\right)_{i,j} =& \Big[\frac{1}{\Delta x}\Big(k_{i+1/2,j}\frac{h_{i+1,j}-h_{i,j}}{\Delta x} - k_{i-1/2,j}\frac{h_{i,j}-h_{i-1,j}}{\Delta x}\Big)\\ &+ \frac{1}{\Delta y}\Big(k_{i,j+1/2}\frac{h_{i,j+1}-h_{i,j}}{\Delta y} - k_{i,j-1/2}\frac{h_{i,j}-h_{i,j-1}}{\Delta y}\Big)\\ &- \alpha\, h_{i,j}^r + P_{i,j}\end{aligned} \tag{1.34a}$$

Et la discrétisation totale est donnée par :

$$
\begin{aligned}
h_{i,j}^{n+1} =& h_{i,j}^{n} + \Delta t \Big[\frac{1}{\Delta x} \Big(k_{i+1/2,j} \frac{h_{i+1,j}^{n} - h_{i,j}^{n}}{\Delta x} - k_{i-1/2,j} \frac{h_{i,j}^{n} - h_{i-1,j}^{n}}{\Delta x} \Big) \\
&+ \frac{1}{\Delta y} \Big(k_{i,j+1/2} \frac{h_{i,j+1}^{n} - h_{i,j}^{n}}{\Delta y} - k_{i,j-1/2} \frac{h_{i,j}^{n} - h_{i,j-1}^{n}}{\Delta y} \Big) \\
&- \alpha \, h_{i,j}^{n} + P_{i,j}^{n} \Big]; \ \forall (i,j) \in \mathbf{F} \text{ et } 0 \leq n \leq N, \qquad (1.35a)
\end{aligned}
$$

$$
(h_{i,j}^{r})^{n} = 0; \ \forall (i,j) \in \mathbf{B} \cup \mathbf{L} \text{ et } 0 \leq n \leq N, \qquad (1.35b)
$$

$$
h_{i,j}^{0} = h_{i,j}^{0}; \ \forall (i,j) \in \mathbf{F}. \qquad (1.35c)
$$

où nous avons introduit d'abord un temps maximum T et un temps d'intervalle $[0, T]$. De même nous avons considéré Δt comme le pas de discrétisation temporelle définie par $\Delta t = T/N$ où $N+1$ est le nombre total de points de la grille. La notation h^n représente l'approximation de $h(\mathbf{x}, t_n)$ pour tout $\mathbf{x} \in \tilde{\Omega}$ et $t_n = n\Delta t$, $0 \leq n \leq N$.

.

Simulation numérique

Nous donnons les résultats numériques d'un exemple dévéloppé dans [31]. Le lecteur pourra consulter cet article pour plus d'informations et d'exemples numériques sur le processus d'assèchement. Dans le suite, la valeur de l'erreur relative est calculée selon la formule ci-après

$$
E_r = \frac{\|H_h - H\|^2}{\|H\|^2} \qquad (1.36)
$$

et l'ordre de convergence est mesuré par

$$
s = \frac{log(\frac{E_h}{E_{2h}})}{log(2)}
$$

où H_h est la solution approchée obtenue par (1.35) et H est la solution exacte du problème. Notre objectif est de rechercher une solution approchée du problème et

d'évaluer l'erreur commise.

Pour ce faire, nous avons considéré le modèle d'un lac tel que la figure 1.1(A droite). Le temps maximal de l'expérience est $T = 10$, les paramètres k, α et r sont supposés connus et donnés ci-après :

$$k(x) = 0.3, \qquad \alpha = 2, \qquad q = 0.8,$$
$$\text{et } h0(x, y) = xy(1 - x)(1 - y).$$

Après simulation du schéma numérique, nous présentons les résultats graphiques dans le domaine considéré aux temps $t = 0; 2; 4; 6; 8$ et 10 (Figure 1.15). La valeur de la hauteur de l'eau est exprimée en couleur. Ces résultats montrent la diminution du niveau de l'eau dans le domaine. Ceci traduit le processus d'assèchement décrit par le modèle (1.32).

1.5 Conclusion

Au terme de ce chapitre, nous avons développé des modèles mathématiques d'une part de l'étude de l'écoulement de la pollution dans une eau de surface et d'autre part du processus d'assèchement d'une étendue d'eau dans un bassin. La nature complexe des domaines sur lesquels sont définis ces problèmes nous a motivé à mettre en place de nouvelles techniques pour le maillage desdits domaines avec une meilleure prise en compte de leur circonférence. De plus nous avons développé des schémas numériques basés d'une part sur une méthode pseudo-spectrale et d'autre part sur une méthode itérative dans la recherche d'une meilleure solution approchée des équations aux dérivées partielles modélisant les problèmes cités plus haut. Les résultats de ces travaux ont abouti à la publications des articles scientifiques [31] (2013), [46](2017) et [47]. Les difficultés rencontrées dans ce travail résident essentiellement dans la collecte des données fiables pour la validation des modèles. nous espérons que des équipes de travail mixte, enseignants-chercheurs et praticiens pourront corriger cette

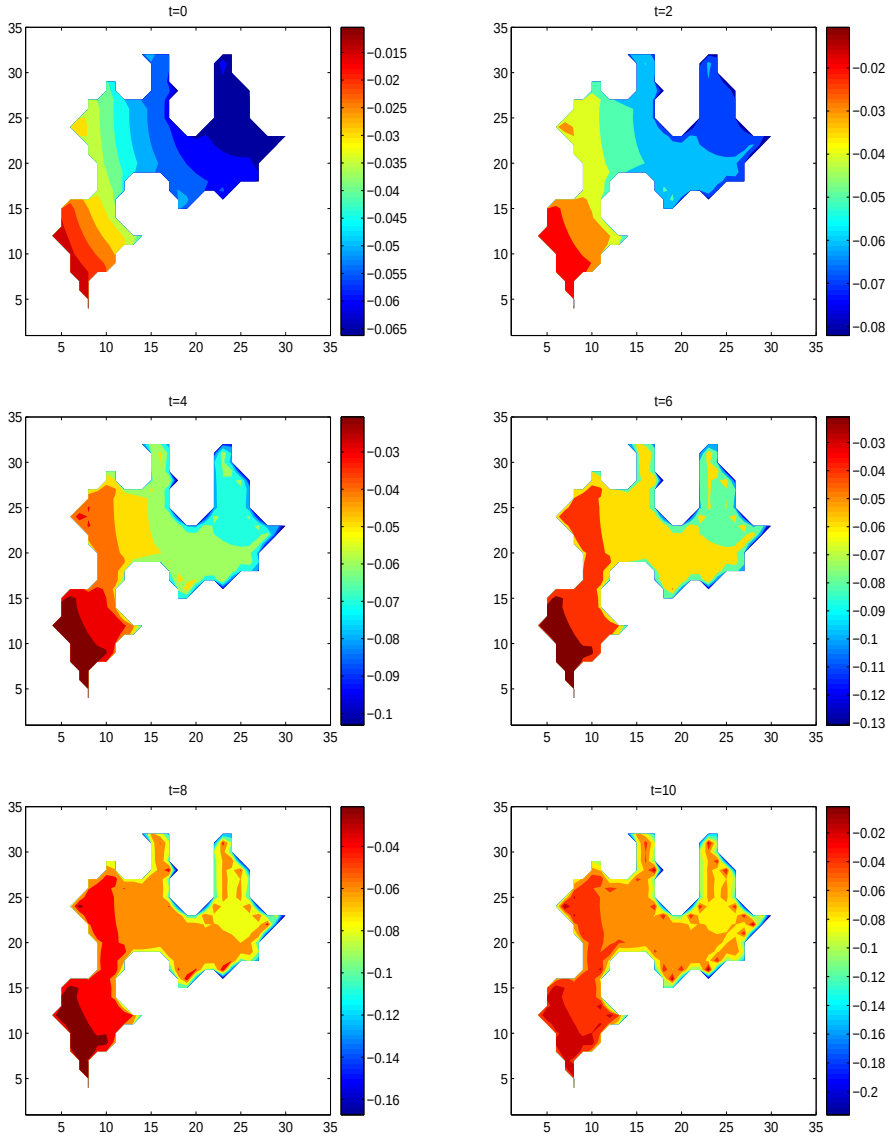


Figure 1.15 – *Les solutions numériques obtenus par (1.35) sur un domaine complexe aux temps : $t = 0; 2; 4; 6; 8; 10$.*

difficulté.

Chapitre 2

Modélisation et simulation numérique d'un problème épidémiologie : Les modèles dépendants de l'âge de la population. [9, 36]

Sommaire

Ce chapitre a fait l'objet de deux publications scientifiques : l'une propose un schéma numérique dans la résolution des modèles dépendants de l'âge de la population et le second fait une analyse asymptotique d'un modèle mathématique

de l'hépatite B. Ces deux papiers sont consultables en annexes 2 et 3.

2.1 Résolution numérique d'un système d'équations non linéaires dépendant de l'âge de la population [9]

Nous présenterons d'abord la méthode numérique et ensuite une illustration sur un modèle donné.

2.1.1 Présentation de la méthode de Heun

Dans ce sous chapitre, considérons le modèle suivant :

$$u_t + u_a = -f\left(a, P(t)\right) u,\ 0 < t \leq T,\ 0 < a < +\infty, \tag{2.1a}$$

$$u\left(t, 0\right) = g\left(t, \int_0^{+\infty} \beta(a, P(t)) u\, da\right), \tag{2.1b}$$

$$u\left(0, a\right) = u^0(a), \tag{2.1c}$$

où

$$P(t) = \int_0^{+\infty} u(t, a) da. \tag{2.2}$$

Les variables indépendantes t et a représentent l'âge et le temps, respectivement. La valeur $u(t, a)$ est la densité de la population au temps t avec un âge a, $f(a, P(t))$ est la fonction de mortalité ou de disparition de la population étudiée en fonction de l'âge, la fonction de natalité ou de fertilité est donnée par $\beta(a, P(t))$. Nous observons que les fonctions f et β dépendent de l'effectif de la population totale $P(t)$ au temps t comme défini dans (2.2). Les conditions initiales et aux bords sont données par u^0 et g, respectivement.

Une étude intensive des modèles linéaires et non-linéaires des populations structurées en âge, en particulier du système (2.1), peut être consultée dans les travaux

de Iannelli ([27], [28]) et Webb([52]). En effet, un grand nombre de méthodes numériques a été développé dans la littérature pour la résolution du système (2.1) durant ces vingt-cinq dernières années.

Pour une excellente review bibliographique de ces méthodes voir les travaux [3] et [7] où une analyse numérique des schémas développés (consistance, stabilité, existence et convergence) est basée sur les travaux de Lòpez-Marcos and Sanz-Serna [35]. Nous mentionnons que notre approche est proche de ces deux travaux. Le premier est proposé par [3], où des methodes de Rung-Kutta ont été utilisées, cependant, l'algorithme des précédents est limité à une durée de vie finie et necessite un grand ordre de regularité sur les hypothèses. Le second est présenté dans [7], il résout un problème sur un support fini mais le fait d'avoir un âge maximum fini impose à la fonction de mortalité d'être non bornée. Ce présent travail résout aussi bien les problèmes définis sur des intervalles bornés ou non en âge.

Pour éviter des temps de calcul élevés et de grandes régularités sur les hypothèses, nous vous proposons une version de la méthode de Heun de second ordre.

2.1.2 Schéma numérique proposé

Conditions de régularités. Pour introduire cette partie, nous rappelons que l'étude de la convergence de tout algorithme itératif et toute approximation polynômiale a besoin d'une certaine régularité des fonctions $u(t,a)$ et f, g, β. Dans la suite, nous assumons les hypothèses suivantes :

$$(H_1)\quad \begin{cases} u_0(a) \text{ est bornée, non négative et} \\ \text{Il existe un âge maximal A tel que} \\ u_0(a) = 0 \text{ avec } a > A; \end{cases}$$

$$(H_2)\quad \begin{cases} f, \beta \in C^3([0,\infty] \times [0,\ +\infty)), \\ \beta(.,P),\ f(.,P) \in C^1([0,\infty), L_\infty([0,\infty)), \\ \beta_P(.,P),\ f_P(.,P) \in C^1([0,\infty), L_\infty([0,\infty)), \\ \beta(.,P),\ f(.,P) \text{ sont à supports compacts dans } [0,\infty), \\ \text{il existe une constante positive } C \text{ tel que} \\ 0 \leq \beta(a,P) \leq C; \end{cases}$$

$$(H_3)\quad \begin{cases} g \in C^1([0,\ T] \times \mathcal{D}_2), \\ \text{où } \mathcal{D}_2 \text{ est un voisinage compact de} \\ \left\{\int_0^{+\infty} \beta(a,P(t))u\,da,\ 0 \leq t \leq T\right\}; \end{cases}$$

$$(H_4)\quad \begin{cases} u^0(0) = g(0,z^0), \\ \text{où } z^0 = \int_0^{+\infty} \beta(a,P^0)u^0(a)\,da \\ \text{et}\ \ P^0 = \int_0^{\infty} u^0(a)da; \end{cases}$$

$$(H_5)\quad \begin{cases} u_a^0(0) = -\Big[f(0,P^0) + \beta(0,P^0)\Big]u^0(0) \\ \qquad +g_t(0,z^0) - g_z(0,z^0)\Big[\int_0^{\infty}\{\beta_a(a,P^0) \\ \qquad +\beta_P(a,P^0)P_t^0 - \beta(a,P^0)f(a,P^0)\}u^0da\Big], \\ \text{où } P_t^0 = u^0(0) - \int_0^{\infty} f(a,P^0)u^0(a)da. \end{cases}$$

Les hypothèses (H_1)-(H_3) garantissent l'existence d'une unique solution $u(t,a)$, tandis que les conditions de compatibilité (H_4)-(H_5) assurent que la solution est

au moins une fois continument dérivable. Nous nous référons à [23] où le théorème suivant est prouvé :

Théorème 2.1 *Supposons que (H_1)-(H_3) soient vérifiés, alors le problème (2.1) a une unique solution non négative $u(t,a)$, globale en temps et $u \in C^1\Big([0,T) \times [0,\infty)/\{(t,t)|t \geq 0\}\Big)$. Si en plus (H_4)-(H_5) sont satisfaites alors $u \in C^1([0,T) \times [0,+\infty))$.*

Pour assurer que $u \in C^k([0,T) \times [0,\infty))$, $k \geq 2$, en plus de (H_1)-(H_5) l'on doit supposer

$$(H_6) \quad \lim_{a\to 0} \frac{d^l}{da^l} u^0(a) = \lim_{t\to 0} \frac{d^l}{dt^l} g(t, P(t)),\ 2 \leq l \leq k.$$

Condition de compatibilité (H_6) joue un rôle essentiel dans la preuve de la consistance du schéma numérique.

Méthode numérique. Le long de toutes les lignes caractéristiques $a_c(t) = t + c$ (c désigne l'âge d'un individu à l'instant $t = 0$) l'équation (2.1a) prend la forme

$$\begin{array}{rl} \frac{d}{dt} u\Big(t, a_c(t)\Big) = & -f\Big(a_c(t), P(t)\Big) u\Big(t, a_c(t)\Big), \\ u(0,c) = & u^0(c). \end{array} \tag{2.3}$$

L'idée principale de cette méthode est de remplacer(2.1) par un système d'équations différentielles ordinaires (2.3) et ensuite appliquer une méthode de Runge-Kutta. Dans la suite nous présentons l'essentiel de cette technique.

Premièrement, nous introduisons des grilles de calcul. Pour un intervalle de temps donné $[0,T]$, $T > 0$, et un nombre total de pas de discrétisation N, nous posons $h = \frac{T}{N}$. L'hypothèse (H_1) implique qu'il existe un entier positif L, tel que u^0 est à support compact dans $[0, Lh]$. En utilisant h, N et L nous définissons :

la grille sur l'axe des temps

$$\mathcal{T} = \{t^n : t^n = nh,\, 0 \leq n \leq N\},$$

la grille sur l'axe des âges

$$\mathcal{A} = \{a_j : a_j = jh,\, j \geq 0\}.$$

En particulier, la grille sur l'axe des âges à l'instant t^n

$$\mathcal{A}^n = \{a_j : a_j = jh,\, 0 \leq j \leq L + n\}$$

et enfin les grilles sur les lignes caractéristiques

$$\mathcal{C} = \{(t^n, a_j) : 0 \leq n \leq N,\, 0 \leq j \leq L + n\}.$$

Nous notons que les points (t^n, a_j) et $(t^n + mh, a_j + mh)$, $m \geq 0$, appartiennent à la même ligne caractéristique.

Ensuite, nous considérons que l'indice j se réfère à l'âge a_j et l'exposant n à l'instant t^n. En utilisant ces notations nous définissons le vecteur

$$U^n = (U_0^n, \ldots, U_{L+n}^n)',$$

dont les composants U_j^n représentent l'approximation numérique de $u(t^n, a_j)$. Nous notons que la taille de U^n dépend du nombre de pas du temps n.

Enfin, nous supposons que U^n est donné et permet d'évaluer U^{n+1} en deux itérations. Dans la première itération, nous calculons les composantes U_j^{n+1}, $1 \leq j \leq L + n + 1$, en intégrant (2.3) dans l'intervalle $[t^n, t^{n+1}]$. A ce niveau, on peut appliquer une méthode de Runge-Kutta sans distinction. Cependant, pour passer d'une itération à une autre dans la détermination des composantes de la solution, l'on

doit faire plusieurs autres calculs dans l'approximation des quadratures. Ceci doit être coûteux puisque chaque niveau d'itération à besoin de l'approximation des intégrales $P(t)$. Dans la seconde itération, nous déterminons la valeur au bord U_0^{n+1} en utilisant (2.1b).

Le temps de calcul est réduit de manière significative si les étapes de calcul (suivant le temps) coïncident avec la grille de calcul (suivant les âges).

Dans les deux itérations ci-dessus, nous devons approximer $P(t^n)$ et $P(t^{n+1})$. La technique la plus simple de faire cela est d'appliquer une formule de quadrature qui prend en compte les points des grilles suivant l'axe des âges prédéfinis, $\mathcal{A}^n$. Dans notre cas, la règle de Simpson est la mieux indiquée. Pour en savoir plus sur les formules de quadrature dans l'approximation des intégrales comme (2.1)) nous invitons le lecteur à consulter [43].

Ainsi l'algorithme se présente comme suit :

$$Q^n = \sum_{j=0}^{L+n} b_j^n U_j^n, \tag{2.4a}$$

$$F_{1,j}^n = -f(a_j, Q^n)U_j^n,\ 0 \le j \le L+n, \tag{2.4b}$$

$$Q^{n+1} = \sum_{j=0}^{L+n+1} b_j^{n+1} U_j^{n+1}, \tag{2.4c}$$

$$F_{2,j}^n = -f(a_{j+1}, Q^{n+1})\Big(U_j^n + hF_{1,j}^n\Big),\ 0 \le j \le L+n, \tag{2.4d}$$

$$U_{j+1}^{n+1} = U_j^n + \tfrac{h}{2}\Big(F_{1,j}^n + F_{2,j}^n\Big),\ 0 \le j \le L+n, \tag{2.4e}$$

$$\hat{Q}^{n+1} = \sum_{j=0}^{L+n+1} b_j^{n+1}\beta(a_j, Q^{n+1})U_j^{n+1}, \tag{2.4f}$$

$$U_0^{n+1} = g(t^{n+1}, \hat{Q}^{n+1}), \tag{2.4g}$$

où les poids de quadrature b_j^n sont donnés par

$$b_j^n = \begin{cases} \frac{h}{3} & \text{si } j = 0 \text{ où } j = L + n; \\ \frac{2h}{3} & \text{si } j \text{ est impaire;} \\ \frac{4h}{3} & \text{autre.} \end{cases}$$

Notons que les équations (2.4a) et (2.4b) sont explicites, tandis que tous les autres sont implicites. Dans la pratique, le système (2.4c)-(2.4g) est résolu par des itérations successives.

2.1.3 Résultats numériques

Nous donnons dans ce paragraphe des exemples d'application de notre schéma numérique. Dans le premier exemple, nous mesurons l'efficacité du schéma numérique (2.4) en terme d'erreur globale, de temps de calcul de la machine et de l'ordre de convergence. Notons que dans cet exemple l'âge maximal est fini. Dans le second exemple nous comparons les solutions numériques obtenues par (2.4) et par la méthode explicite de Euler d'ordre deux (Une fois encore l'âge maximum est fini également). Le dernier exemple traite le cas d'un problème avec un maximum d'âge illimité. Dans tous ces exemples l'erreur globale est calculée par la formule suivante

$$E_h = \max_{0 \leq n \leq N} \|U_h^n - U^n\|_n. \tag{2.5}$$

Exemple 1 Dans cet exemple le maximum d'âge A est fixé à 1 unité. La fonction de mortalité, l'âge spécifique de fertilité, la densité initiale et la fonction de naissance

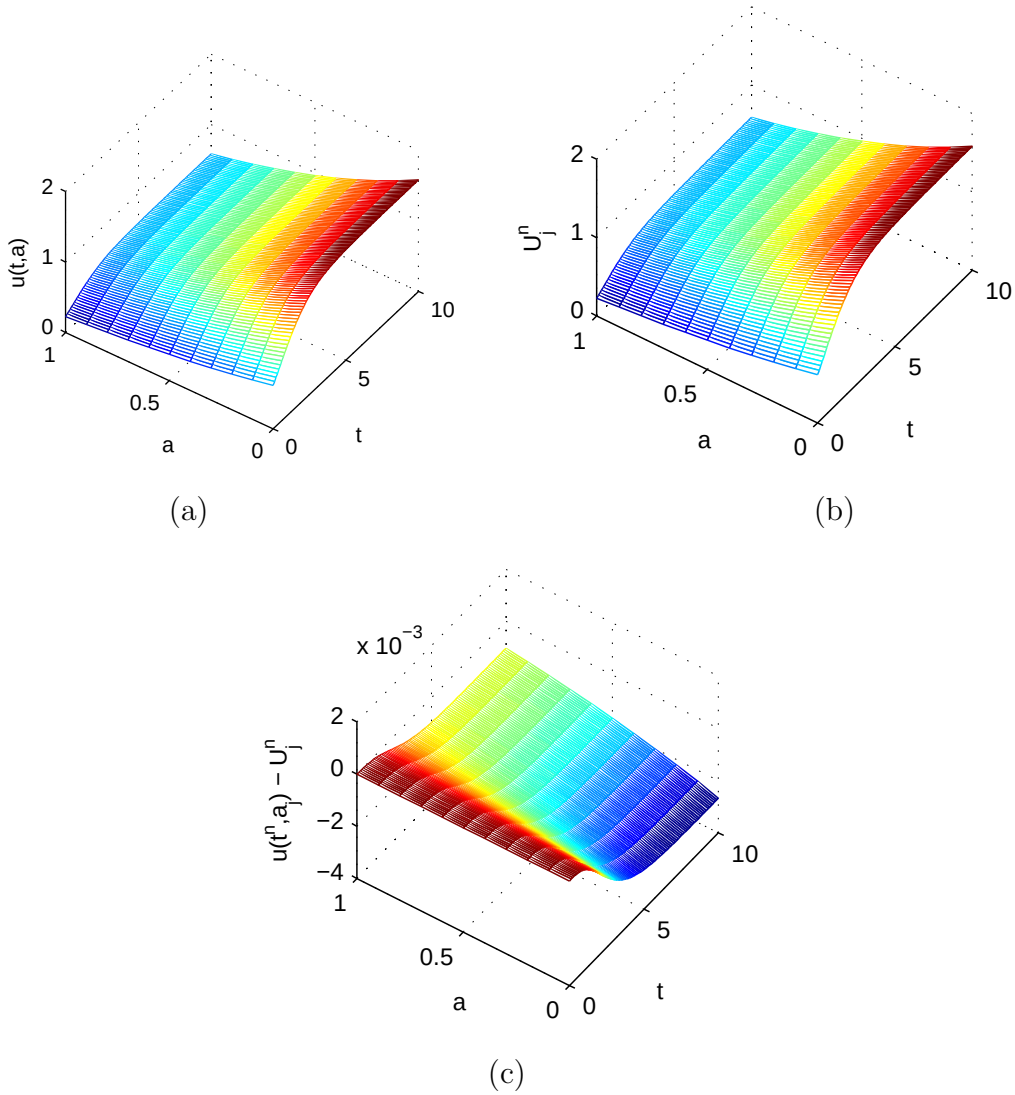


Figure 2.1 – *La solution exacte(a), la solution approchée(b) et l'erreur(c) pour N=100.*

Tableau 2.1 – Erreurs, temps machine (secondes), et l'ordre de convergence de l'exemple 1.

N	20	40	80	160	320
E_h	$0.534 \cdot 10^{-5}$	$0.212 \cdot 10^{-5}$	$0.325 \cdot 10^{-6}$	$0.796 \cdot 10^{-7}$	$0.195 \cdot 10^{-7}$
CPU	0.094	0.187	0.515	1.497	6.74
ordre		2.049	1.988	2.030	2.027

sont données par :

$$f(a,z) = -z, \quad \beta(a,z) = \frac{aze^{-a}}{(1+z)^2}, \quad u^0(a) = \frac{e^{-a}}{2-e^{-A}},$$
$$g(z,t) = \frac{4z(2-2e^{-A}+e^{-t})^2}{(1-e^{-A})(1-(1+2A)e^{-2A})(1-e^{-A}+e^{-t})},$$

respectivement. La solution exacte de ce problème est (voir [3])

$$u(t,a) = \frac{e^{-a}}{1-e^{-A}+e^{-t}}.$$

Les résultats de la simulation numérique (dans l'intervalle de temps $[0,10]$) sont présentés dans le tableau 2.1.

Notons aussi que $u(t,a) \in C^\infty([0,T) \times [0,1))$ et le théorème 1 prédisent que $E_h = \mathcal{O}(h^2)$. La dernière colonne du tableau 2.1 est en accord avec la théorie. Les solutions exactes et numériques et les erreurs sont présentées dans Figure 2.1.

Exemple 2 Ici nous comparons des solutions obtenues en utilisant le schéma numérique (2.4) et le schéma de Euler explicite. Nous utilisons les données : $A = 0.9$, $T = 1$,

$$f(a,z) = -\frac{1}{1-a} - z, \quad \beta(a,z) = 4,$$
$$g(t,z) = z, \quad u^0(a) = 4(1-a)e^{-\alpha a}.$$

Tableau 2.2 – Erreurs et Temps machine (CPU) (en secondes) pour l'exemple 3.

	N	20	40	60	80
$L = 5N$	E_h	$0.121 \cdot 10^{-1}$	$0.233 \cdot 10^{-1}$	$0.36 \cdot 10^{-1}$	$0.482 \cdot 10^{-1}$
	CPU	0.109	0.20	0.327	0.499
$L = 10N$	E_h	$0.816 \cdot 10^{-5}$	$0.421 \cdot 10^{-5}$	$0.306 \cdot 10^{-5}$	$0.124 \cdot 10^{-5}$
	CPU	0.14	0.29	0.515	0.827
$L = 1000N$	E_h	$0.782 \cdot 10^{-5}$	$0.335 \cdot 10^{-5}$	$0.294 \cdot 10^{-5}$	$0.169 \cdot 10^{-5}$
	CPU	3.1	10.28	22.94	42.073
$L = 2000N$	E_h	$0.782 \cdot 10^{-5}$	$0.335 \cdot 10^{-5}$	$0.294 \cdot 10^{-5}$	$0.169 \cdot 10^{-5}$
	CPU	4.618	17.082	38.204	70.184

La solutions exacte est donnée par

$$u(a,t) = \frac{4\alpha(1-a)e^{-\alpha a}}{(\alpha-1)e^{-\alpha t}+1},$$

où $\alpha = 2.5569290855$.

Les solutions numériques obtenues par la méthode de Euler explicite et par les schéma numérique (2.4) sont présentées dans Figure 2.2.

Notons que les solutions numériques à faible ordre de convergence développent des sauts de discontinuités le long de la ligne caractéristique $a = t$. L'amplitude de ces sauts diminue lorsque $N \to \infty$ mais ne disparait pas complètement. La situation est complètement différente si nous utilisons (2.4), les solutions numériques restent continues pour tout N.

Exemple 3 Dans ce exemple $A = \infty$, l'âge spécifique de fertilité et la fonction de mortalité sont les mêmes que dans l'exemple 1, la fonction de naissance et la fonction initiale sont données par :

$$g(t,z) = \frac{4z(2+e^{-t})^2}{(1+e^{-t})}, \quad u^0(a) = \frac{1}{2}e^{-a},$$

respectivement. La solution exacte est

$$u(t,a) = \frac{e^{-a}}{1+e^{-t}}.$$

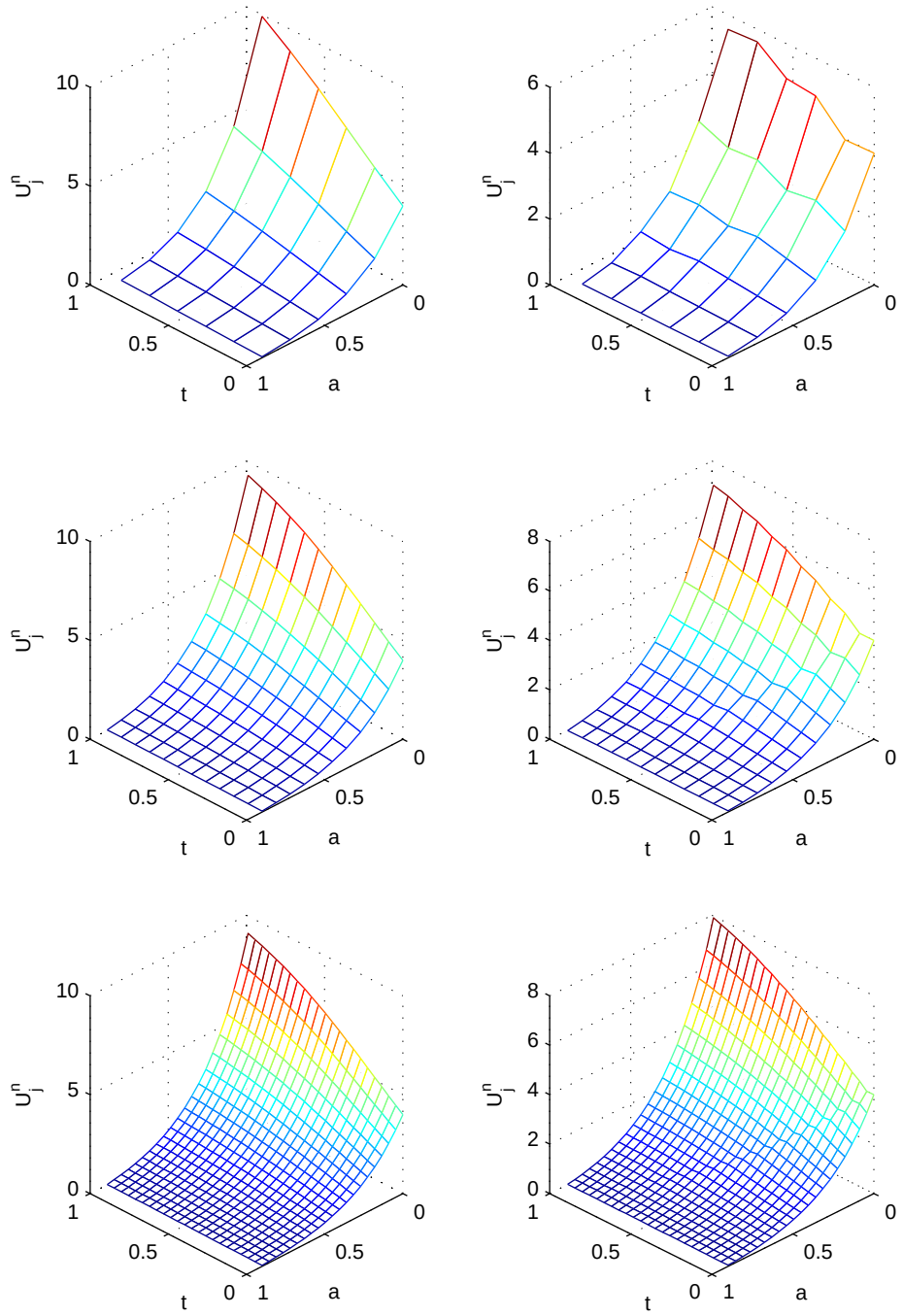


Figure 2.2 – (Colonne à gauche)Les solutions numériques obtenues par (2.4) et (colonne à droite) par Euler explicite pour $N = 5;\ 10 et\ 20$.

2.1 Résolution numérique d'un système d'équations non linéaires dépendant de l'âge de la population [9]

Pour des raisons de simulation numérique, l'infini sera matérialisé par la plus grande valeur que nous souhaitons. A cet effet, nous prenons L dans ($5N \leq L \leq 2000N$). Il est claire que le temps de calcul augmentera en fonction de la taille de L. Cependant, la théorie développée dans la section 3 atteste que $E_h = \mathcal{O}(h^{\min\{k-1,2\}})$ dès que $\operatorname{supp} u^0 \subset [0, hL]$. Aussi dans la pratique, il n'y a pas besoin d 'accroitre la valeur de L indéfiniment. Les valeurs des calculs resteront constantes pour tout $L \geq L^*$, où L^* est le plus petit entier qui satisfait $\operatorname{supp} u^0 \subset [0, hL^*]$. Les résultats des simulations (dans l'intervalle de temps $[0, 10]$) présentés dans Table 2.2 confirment complètement cette simple observation.

2.2 Exemple de cas : Étude asymptotique de la dynamique de transmission de l'hépatite B dépendant de l'âge de la population [36].

Au cours des études des polluants dans les eaux de surface (les chapitres précédents), nous avons énuméré des micros organismes provenant des eaux usées notamment des virus. Ce qui fait de l'eau polluée un vecteur potentiel de transmission des maladies hydriques. Nous nous intéressons dans ce chapitre à la modélisation et à l'étude asymptotique de la transmission de l'hépatite B.

2.2.1 Mise en équation du problème

Le virus de l'hépatite B(VHB) se transmet par les sécrétions corporelles comme le sang, la salive et les sécrétions vaginales. Dans ce paragraphe nous partons du modèle de la transmission de VHB introduit dans [55] auquel nous avons fait des hypothèses réalistes. D'abord la population est divisée en six sous-classes que sont :

1-Susceptibles : les individus sains et exposés à la maladie ;

2-Latents : les individus infectés, mais pas encore infectieux ;

3-Infectés : les individus infectés et malade ;

4-Chroniques : les personnes infectées en phase aigüe ;

5-Rétablis ou Guéris : les individus qui se sont rétablis de l'infection et qui sont maintenant des individus immunisés ;

6-Vaccinés : ceux immunisés par le vaccin.

Ensuite nous supposerons que :

(1) Les phases latents, infectés et chroniques sont séparées. Seuls les individus en phase infectés et chronique sont infectieux.

(2) Tous les individus en phase latente développent d'abord une phase aiguë

(3) Certains individus ayant une infection aiguë progressent vers l'état chronique et plus tard développent l'immunité tandis que d'autres développent l'immunité sans progresser vers l'état chronique.

(4) Il existe une possibilité de traitement (ou de récupération) pendant la phase aiguë de l'infection et l'état de porteur chronique de l'infection.

(5) Il n'y a pas de proportion d'infectés périnatales de nouveau-nés (des mères porteuses)

(6) il y a une proportion de naissances avec une vaccination réussie et une proportion de nouveau-nés susceptibles

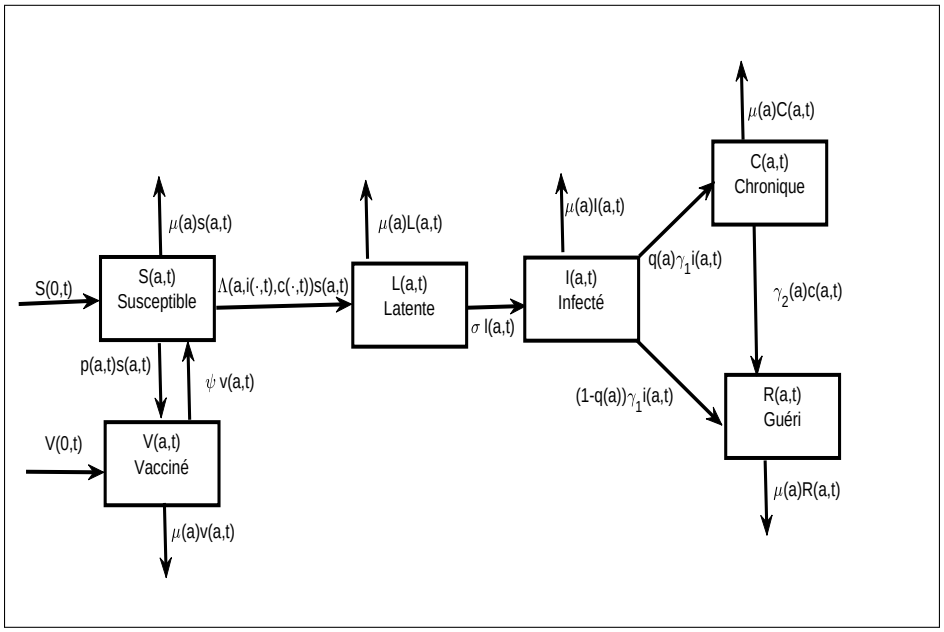


Figure 2.3 – *Le diagramme du flux de transmission de la maladie.*

Dans la suite nous désignons par $s(a,t)$, $l(a,t)$, $i(a,t)$, $c(a,t)$, $r(a,t)$ et $v(a,t)$ les fonctions de densités correspondantes pour ces classes épidémiologiques dépendant de l'âge et du temps respectivement.

Le choix de la force d'infection est determinant dans le processus de modélisation de l'évolution de la maladie, à cet effet nous considérons une forme constitutive

2.2 Exemple de cas : Étude asymptotique de la dynamique de transmission de l'hépatite B dépendant de l'âge de la population [36].

Tableau 2.3 – Définition des paramètres de la modélisation

Paramètres	Description
$\beta(a)$	Fonction d'âge spécifique de fertilité de la population
$\mu(a)$	Taux de mortalité naturel de la population
σ	Taux d'individus se déplaçant des latents vers la phase aigüe
γ_1	Taux d'individu se déplaçant vers les porteurs chroniques
$\gamma_2(a)$	Taux d 'individus se déplaçant des porteurs chroniques vers les vaccinés
$p(a,t)$	Taux de vaccination contre VHB
ψ	Taux de l'immunité induite par la vaccination
$(1-\omega)$	Proportion de naissances avec des vaccinations à la naissance
$q(a)$	Probabilité qu'un individu en phase aigu passe en phase chronique
Λ	Taux d'infection (ou force d'infection)

intercohort séparable suivante introduit par [17] :

$$\Lambda(a, i(\cdot,t), c(\cdot,t)) = k(a) \int_0^{a_+} [h_1(a)i(a,t) + h_2(a)c(a,t)]\, da \tag{2.6}$$

où $h_1(a)$ et $h_2(a)$ sont des taux d'infection d'âge spécifique correspondant à la phase aïgue et chronique respectivement. $k(a)$ est le taux de contagion de l'âge spécifique, ici nous assumons que les deux phases ont le même taux de contagion, a_+ est l'âge maximal d'un individu et que $h_1(a)$, $h_2(a)$ et $k(a)$ vérifient les conditions suivantes :

$$h_1, h_2, k \in \mathrm{L}^\infty([0, a_+]), \quad h_1(a), h_2(a), k(a) \geq 0 \quad . \tag{2.7}$$

Enfin, la dynamique du modèle épidémiologique structuré par âge pour la transmission du VHB, schématisée par la figure 2.3 peut être décrite par le problème

suivante :

$$\begin{aligned}
\partial_t s(a,t) &= -\partial_a s(a,t) - \mu(a)s(a,t) + \psi v(a,t) - s(a,t)\Lambda(a, i(\cdot,t), c(\cdot,t)) \\
&\quad - p(a,t)s(a,t), \\
\partial_t l(a,t) &= -\partial_a l(a,t) - \mu(a)l(a,t) - \sigma l(a,t) + s(a,t)\Lambda(a, i(\cdot,t), c(\cdot,t)), \\
\partial_t i(a,t) &= -\partial_a i(a,t) - \mu(a)i(a,t) + \sigma l(a,t) - \gamma_1 i(a,t), \\
\partial_t c(a,t) &= -\partial_a c(a,t) - \mu(a)c(a,t) - \gamma_2(a)c(a,t) + q(a)\gamma_1 i(a,t), \\
\partial_t r(a,t) &= -\partial_a r(a,t) - \mu(a)r(a,t) + \gamma_2(a)c(a,t) + (1-q(a))\gamma_1 i(a,t), \\
\partial_t v(a,t) &= -\partial_a v(a,t) - \mu(a)v(a,t) - \psi v(a,t) + p(a,t)s(a,t),
\end{aligned} \tag{2.8}$$

avec des conditions aux bords (les nouveaux-nés)

$$\begin{aligned}
s(0,t) &= \omega \int_0^{a_+} \beta(a) \left[s(a,t) + l(a,t) + i(a,t) + r(a,t) + v(a,t) + c(a,t) \right] da, \\
l(0,t) &= i(0,t) = c(0,t) = r(0,t) = 0, \\
v(0,t) &= (1-\omega) \int_0^{a_+} \left[s(a,t) + l(a,t) + i(a,t) + r(a,t) + v(a,t) + c(a,t) \right] da,
\end{aligned} \tag{2.9}$$

et les conditions initiales

$$\begin{aligned}
s(a,0) &= \overset{o}{s}(a), \quad l(a,0) = \overset{o}{l}(a), \quad i(a,0) = \overset{o}{i}(a), \\
c(a,0) &= \overset{o}{c}(a), \quad r(a,0) = \overset{o}{r}(a), \quad v(a,0) = \overset{o}{v}(a).
\end{aligned} \tag{2.10}$$

Les définitions des paramètres utilisés dans la modélisation peuvent être consultées dans le tableau 2.3.

2.2.2 Équations singulièrement perturbées

Dans la construction du modèle épidémiologique dépendant de l'âge (2.8-2.10), nous avons pris en compte à la fois la dynamique démographique et l'infection.

2.2 Exemple de cas : Étude asymptotique de la dynamique de transmission de l'hépatite B dépendant de l'âge de la population [36].

Fort malheureusement le processus de contamination agit sur une échelle de temps différente de celle du processus démographique.

A titre d'illustration, en se basant sur les résultats numériques obtenus pour les paramètres dans (2.1-2.10) provenant de [54, 55] : les taux de mortalité et de natalité sont mesurés en unités 1/70 ans, où 70 est considéré comme la durée de vie moyenne de la population étant donné que nous étudions une population humaine. De plus, nous notons que $\beta = 0.0121$, $\mu = 0.00693$, $\mu_1 = 0.002$, $\psi = 0.1$, $\Lambda \approx 0.16$, $\sigma = 6/\text{an}$, $\gamma_1 = 4/\text{an}$ et $\gamma_2 = 0.025/\text{an}$. Nous faisons le constat suivant : en utilisant 70 ans comme unité de temps dans le modèle(2.8-2.10), les valeurs numériques pour σ, γ_1 et γ_2 seraient multipliées par 70 et ce résultat donne $\sigma = 420$, $\gamma_1 = 280$ et $\gamma_2 = 1.75$. Ceci montre que les processus induits par σ et γ_1 (processus de contamination) sont plus rapides que ceux induits par β, μ (processus démographique), ψ et Λ ; et le processus induit par γ_2 est légèrement plus rapide que ceux induits par β, μ, ψ et Λ. Dans le modèle asymptotique nous considérons la même échelle de temps aussi bien pour les durées des processus d'infection que démographique, de telle sorte que les processus d'infection induis par les paramètres σ, γ_1 et γ_2 sont plus rapides que ceux induits par les paramètres β, μ (processus démographique), ψ et Λ. Dans la suite, nous considérons ϵ un petit paramètre reflétant le rapport des échelles de temps typiques des processus vitaux et épidémiologiques. Ainsi, le modèle asymptotique de (2.8-2.10) se résume comme suit :

$$\begin{cases} \partial_t \mathbf{u}_\epsilon = \mathcal{S}\mathbf{u}_\epsilon + \mathcal{M}\mathbf{u}_\epsilon + \mathcal{F}(\mathbf{u}_\epsilon) + \dfrac{1}{\epsilon}\mathcal{C}\mathbf{u}_\epsilon, \\ \mathbf{u}_\epsilon(0,t) = \mathcal{B}\left[\mathbf{u}_\epsilon(\cdot,t)\right], \\ \mathbf{u}_\epsilon(a,0) = \overset{\circ}{\mathbf{u}}, \end{cases} \tag{2.11}$$

où

$$\mathbf{u}_\epsilon = (s_\epsilon, l_\epsilon, i_\epsilon, c_\epsilon, r_\epsilon, v_\epsilon),$$

$$\mathcal{S} = \text{diag}\{-\partial_a, -\partial_a, -\partial_a, -\partial_a, -\partial_a, -\partial_a\}$$

$$\mathcal{M}(a) = \begin{pmatrix} -\mu(a) & 0 & 0 & 0 & 0 & \psi \\ 0 & -\mu(a) & 0 & 0 & 0 & 0 \\ 0 & 0 & -\mu(a) & 0 & 0 & 0 \\ 0 & 0 & 0 & -\mu(a) & 0 & 0 \\ 0 & 0 & 0 & 0 & -\mu(a) & 0 \\ 0 & 0 & 0 & 0 & 0 & -\mu(a) - \psi \end{pmatrix},$$

$$[\mathcal{F}(\mathbf{u})](a) = \begin{pmatrix} -s(a)p(a) - s(a)\Lambda(a,i,c) \\ s(a)\Lambda(a,i,c) \\ 0 \\ 0 \\ 0 \\ s(a)p(a) \end{pmatrix}$$

et

$$[\mathcal{C}\mathbf{u}](a) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\sigma & 0 & 0 & 0 & 0 \\ 0 & \sigma & -\gamma_1 & 0 & 0 & 0 \\ 0 & 0 & q(a)\gamma_1 & -\gamma_2(a) & 0 & 0 \\ 0 & 0 & (1-q(a))\gamma_1 & \gamma_2(a) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \mathbf{u}(a).$$

En outre, l'opérateur borné $\mathcal{B} : \mathrm{L}^1([0, a_+], \mathbb{R}^6) \to \mathbb{R}^6$ est défini par

$$\mathcal{B}\mathbf{u} = \int_0^{a_+} B(a)\mathbf{u}(a)\, da,$$

avec

$$B(a) = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ b_{61} & b_{62} & b_{63} & b_{64} & b_{65} & b_{66} \end{pmatrix},$$

$b_{1k} = \omega\beta(a)$ et $b_{6k} = (1-\omega)\beta(a)$, $k = 1, \ldots, 6$,

Nous visons à étudier le comportement de la solution $\mathbf{u}_\epsilon = (s_\epsilon, l_\epsilon, i_\epsilon, c_\epsilon, r_\epsilon, v_\epsilon)$, de (2.11) lorsque $\epsilon \to 0$.

Taux de reproduction

Étant donnée une maladie, une question fondamentale est de savoir si elle peut se propager dans la population. Ceci revient à calculer le nombre moyen d'individus qu'une personne infectieuse pourra infecter, tant qu'elle sera contagieuse. Ce nombre est appelé le taux de reproduction. Il est considéré dans une population où tous les individus sont sains, sauf l'individu infectieux introduit. Si $R_0 < 1$, alors un individu en infecte en moyenne moins d'un, ce qui signifie que la maladie disparaîtra de la population à terme. À l'opposé, si $R_0 > 1$, alors la maladie peut se propager dans la population. Déterminer R en fonction des paramètres du modèle permet ainsi de calculer les conditions dans lesquelles la maladie se propage. On définit dans la suite le taux de reproduction net par :

$$R_\mu = \int_0^{a_+} \beta(a)\Pi_\mu(a)\, da \tag{2.12}$$

où

$$\Pi_\mu(a) := e^{-\int_0^a \mu(s)\, ds} \tag{2.13}$$

est la probabilité de survie d'un individu jusqu'à l'âge a, et M une constante tel que

$$\|n(t)\| \leq Me^{\lambda_\mu t}\|\overset{\circ}{n}\|. \tag{2.14}$$

En effet si nous supposons que $\mathbf{u}_\epsilon$ vérifie (2.11) alors, en additionnant les équations dans (2.11), il revient que la population totale,

$$n(a,t) = s_\epsilon(a,t) + l_\epsilon(a,t) + i_\epsilon(a,t) + c_\epsilon(a,t) + r_\epsilon(a,t) + v_\epsilon(a,t),$$

satisfait

$$\begin{cases} \partial_t n(a,t) = -\partial_a n(a,t) - \mu(a)n(a,t), \\ n(0,t) = \int_0^{a_+} \beta(a)n(a,t)\,da, \\ n(a,0) = \overset{\circ}{n}(a), \end{cases} \tag{2.15}$$

et est indépendant de ϵ. Il résulte de [8, 27, 45] qu'il existe une valeur propre dominante $\lambda_\mu \leq \overline{\beta} - \underline{\mu}$, définie comme une unique solution de

$$\int_0^{a_+} e^{-\lambda a}\beta(a)\Pi_\mu(a)\,da = 1. \tag{2.16}$$

Lemme 2.1 *λ_μ est négative, zéro où positive si et seulement si le taux de reproduction net, R_μ, est respectivement, plus petit, égal ou supérieur à un.*

Comme nous l'avons mentionné plus tôt, la mise à l'échelle considérée est réaliste pour les modèles dans lesquels il n'y a pas de changements significatifs de la population totale. Par conséquent, tout au long du document, nous supposerons

$$R_\mu \leq 1. \tag{2.17}$$

2.2.3 Quelques résultats théoriques

On admettra dans la suite que $\overline{\beta} \leq \underline{\mu}$. avec $\overline{w}$ la solution du problème de McKendrick-von Foerster

$$\begin{aligned} \partial_t \overline{w}(a,t) &= -\partial_a \overline{w}(a,t) - \mu(a)\overline{w}(a,t), \\ \overline{w}(0,t) &= 0, \\ \overline{w}(a,0) &= \overset{\circ}{w}(a) = \overset{\circ}{l}(a) + \overset{\circ}{i}(a) + \overset{\circ}{c}(a) + \overset{\circ}{r}(a), \end{aligned} \tag{2.18}$$

et $\overline{v}$ est la solution du problème de McKendrick-von Foerster

$$\begin{aligned} \partial_t \overline{v}(a,t) &= -\partial_a \overline{v}(a,t) - \mu(a)\overline{v}(a,t) - \psi\overline{v}(a,t) \\ &\quad - p(a,t)\overline{v}(a,t) + p(a,t)(n(a,t) - \overline{w}(a,t)), \\ \overline{v}(0,t) &= (1-\omega)n(0,t), \\ \overline{v}(a,0) &= \overset{\circ}{v}(a). \end{aligned} \tag{2.19}$$

De plus, nous considérons les espaces suivants : $\mathbf{X} = \mathrm{L}^1([0,a_+],\mathbb{R}^6)$ et $\mathbf{X}_+ = \mathrm{L}^1\left([0,a_+],\mathbb{R}^6_+\right)$. On associe à $\mathbf{X}$ la norme, $\|\cdot\|_X$.

Enfin nous faisons les hypothèses suivantes :

(A1) : $\mu \in \mathrm{L}^1_{\mathrm{loc}}([0,a_+))$, $\int_0^{a_+} \mu(s)\,ds = \infty$ avec $\underline{\mu} > 0$;

(A2) : $\beta \in \mathrm{L}^\infty([0,a_+])$;

(A3) : $q, \gamma_2 \in \mathrm{W}^{1,\infty}([0,a_+])$ avec $\underline{q} > 0$, $\underline{\gamma_2} > 0$;

(A4) : $p \in \mathcal{C}([0,a_+] \times [0,T])$;

(A5) : $a_+ < \infty$ où a_+ est l'âge maximal, ceci signifie qu'aucun individu ne peut vivre indéfiniment ;

(A6) : $\Pi_\mu(a_+) = 0$ avec $\Pi_\mu(a) := e^{-\int_0^a \mu(s)\,ds}$ la probabilité de survie.

Sous les hypothèses citées ci-dessus, il est prouvé dans [13], l'existence d'une solution $\mathbf{u}_\epsilon$ du système (2.11).

La preuve est basée sur le fait que l'opérateur $\mathcal{A} := \mathcal{S} + \mathcal{M}$ défini dans

$$D(\mathcal{A}) = \{\mathbf{u} \in D(\mathcal{S}) \cap D(\mathcal{M});\ \mathbf{u}(0) = \mathcal{B}\mathbf{u}\} \tag{2.20}$$

génère un $\mathcal{C}_0$-semi-groupe positif, noté $\left(e^{t\mathcal{A}}\right)_{t\geq 0}$. Puisque, pour un ϵ fixé, (2.15) est un système linéaire quadratique perturbé.

Théorème 2.2 *Si* $\overset{o}{\mathbf{u}} = (\overset{o}{s}, \overset{o}{l}, \overset{o}{i}, \overset{o}{c}, \overset{o}{r}, \overset{o}{v}) \in \mathbf{X}_+$, *alors il existe une unique solution faible globale lorsque* $t \to \infty$, $\mathbf{u}_\epsilon(t) = (s_\epsilon(t), l_\epsilon(t), i_\epsilon(t), c_\epsilon(t), r_\epsilon(t), v_\epsilon(t)) \in \mathcal{C}([0,\infty), \mathbf{X})$ *à (2.4). Cette solution devient une solution classique si* $\overset{o}{\mathbf{u}} \in D(\mathcal{A})$. *Dans un tel cas nous obtenons, en particulier, que* $\mathbf{u}_\epsilon$ *est continue sur* $[0, a_+] \times [0, T]$ *pour tout* $0 \leq T < \infty$, $\mathbf{u}_\epsilon \in D(\mathcal{S}) \cap D(\mathcal{M})$ *et (2.4) est satisfait presque partout sur* $[0, a_+] \times [0, T]$.

La preuve de ce théorème est à consultée dans [12, 13]. De simple calculs montre que, voir [29], $\left(e^{t\mathcal{A}}\right)_{t\geq 0}$ vérifie l'estimation suivante

$$\|e^{t\mathcal{A}}\| \leq e^{(\overline{\beta}-\underline{\mu})t}. \tag{2.21}$$

Cette estimation peut être améliorée. En fait, tel que nous l'avons mentionné plus haut, si $(s_\epsilon(t), l_\epsilon(t), i_\epsilon(t), , c_\epsilon(t), r_\epsilon(t), v_\epsilon(t))$ est une solution classique de (2.11), alors $n(a,t) = s_\epsilon(a,t)+l_\epsilon(a,t)+i_\epsilon(a,t)+c_\epsilon(a,t)+r_\epsilon(a,t)+v_\epsilon(a,t)$ est une solution classique de (2.15). Nous désignons par $(A, D(A))$ le générateur du semi-groupe $\left(e^{tA}\right)_{t\geq 0}$ pour (2.15), sur un domaine défini comme dans(2.20). De plus, si $\overset{o}{\mathbf{u}} \geq 0$ entraine que $\mathbf{u}_\epsilon(t) = (s_\epsilon(t), l_\epsilon(t), i_\epsilon(t), c_\epsilon(t), r_\epsilon(t), v_\epsilon(t)) \geq 0$. Ainsi chaque composante de $\mathbf{u}_\epsilon$ est majorée par une solution n de (2.15) indépendant de ϵ :

$$\begin{array}{lll} 0 \leq s_\epsilon(a,t) \leq n(a,t), & 0 \leq l_\epsilon(a,t) \leq n(a,t), & 0 \leq i_\epsilon(a,t) \leq n(a,t), \\ 0 \leq c_\epsilon(a,t) \leq n(a,t), & 0 \leq r_\epsilon(a,t) \leq n(a,t) \quad \text{and} & 0 \leq v_\epsilon(a,t) \leq n(a,t) \end{array} \tag{2.22}$$

pour tout $t \geq 0$ et $a \in [0, a_+]$ ainsi l'inégalité (2.14) est satisfaite.

2.2.4 Développement asymptotique du modèle

En se basant sur l'approche générale de l'analyse asymptotique [11, 13], nous cherchons à identifier *l'espace hydrodynamique* V du système d'équations singulièrement perturbées (2.11) qui, dans ce cas, est donné par l'espace nul de $\mathcal{C}$. Dans ce contexte, a est traité comme un paramètre. Après plusieurs calculs et en supposant a comme une constante on obtient :

$$V = \{\mathbf{u} \in \mathbb{R}^6;\ \mathbf{u} = (u_1, 0, 0, 0, u_5, u_6), u_1, u_5, u_6 \in \mathbb{R}\} = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\},$$

où $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ constituent une base de V. L'espace spectral complémentaire W, appelé l'espace cinétique, généré par les valeurs propres $\lambda_4 = -\sigma$, $\lambda_5 = -\gamma_1$, et $\lambda_6 = -\gamma_2(a)$, respectivement associées aux vecteurs définis par :

$$\mathbf{e}_4 = \begin{pmatrix} 0 \\ 1 \\ -\frac{\sigma}{\sigma-\gamma_1} \\ \frac{\sigma q(a)\gamma_1}{(\sigma-\gamma_1)(\sigma-\gamma_2(a))} \\ \frac{\gamma_1}{\sigma-\gamma_1} \cdot \frac{(1-q(a))\sigma-\gamma_2(a)}{\sigma-\gamma_2(a)} \\ 0 \end{pmatrix},\ \mathbf{e}_5 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -\frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)} \\ -\frac{(1-q(a))\gamma_1-\gamma_2(a)}{\gamma_1-\gamma_2(a)} \\ 0 \end{pmatrix} \text{ et } \mathbf{e}_6 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix},$$

respectivement.

Pour achever le décomposition de $\mathbf{u} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3 + c_4\mathbf{e}_4 + c_5\mathbf{e}_5 + c_6\mathbf{e}_6$ dans les espaces hydrodynamique et cinétique, nous devons identifier les coéfficients c_i, $i = 1, \ldots, 6$. En utilisant des résultats standard de l'algèbre linéaire [13], $c_i = \mathbf{f}_i \cdot \mathbf{u}/\mathbf{f}_i \cdot \mathbf{e}_i$ où $\mathbf{f}_i$ sont les vecteurs propres de gauche de $\mathcal{C}$ correspondants aux vecteurs propres de $\mathbf{e}_i$, $i = 1, \ldots, 6$. Ici, le contexte est plus compliqué dans la mesure où V est de dimension trois et nous devons choisir ses bases convenablement. Sachant que $\mathbf{f}_1 = (1, 1, 1, 1, 1, 1)$ est vecteur propre de gauche de $\mathcal{C}$ correspondant à la valeur propre nulle et $\mathbf{f}_1 \cdot \mathbf{u} = u_1 + u_2 + u_3 + u_4 + u_5 + u_6$, il jouera un rôle significatif dans

l'analyse de (2.15).

Ainsi, pour faciliter les calculs, nous prenons $\mathbf{e}_2 = (-1,0,0,0,0,1) \in V$ qui est orthogonal à $\mathbf{f}_1$. En prenant $\mathbf{f}_3 = (0,0,0,0,0,1)$ ceci entraine $\mathbf{e}_1 = (1,0,0,0,0,0) \in V$ et $\mathbf{e}_2 = (-1,0,0,0,1,0) \in V$. Finalement $\mathbf{f}_4 = (0,1,0,0,0,0)$, $\mathbf{f}_5 = (0,\sigma/(\sigma - \gamma_1),1,0,0,0)$, $\mathbf{f}_6 = (0,\sigma q(a)\gamma_1/(\sigma - \gamma_2(a))(\gamma_1 - \gamma_2(a)), q(a)\gamma_1/(\gamma_1 - \gamma_2(a)),1,0,0)$. Donc nous avons la décomposition suivante :

$$\begin{aligned}\mathbf{u} = {} & (u_1+u_2+u_3+u_4+u_5+u_6)\mathbf{e}_1 + (u_2+u_3+u_4+u_5)\mathbf{e}_2 + u_6\mathbf{e}_3 + u_2\mathbf{e}_4 \\ & + \left(\frac{\sigma}{\sigma-\gamma_1}u_2 + u_3\right)\mathbf{e}_5 + \left(\frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\sigma}{\sigma-\gamma_1}u_2 + \frac{q\gamma_1}{\gamma_1-\gamma_2}u_3 + u_4\right)\mathbf{e}_6.\end{aligned} \tag{2.23}$$

Approximation globale

Dans le cadre de l'étude de l'approximation globale de la solution de (2.11) et en se basant sur la décomposition (2.23), nous procédons à la définition de nouvelles variables accumulées que sont :

$$\begin{aligned} n = {} & s_\epsilon + l_\epsilon + i_\epsilon + c_\epsilon + r_\epsilon + v_\epsilon; \\ w_\epsilon = {} & l_\epsilon + i_\epsilon + c_\epsilon + r_\epsilon, \\ z_\epsilon = {} & \tfrac{\sigma}{\sigma-\gamma_1} l_\epsilon + i_\epsilon, \\ x_\epsilon = {} & \tfrac{q\gamma_1}{\gamma_1-\gamma_2} z_\epsilon + c_\epsilon \end{aligned}$$

et laissons inchangées les variables l_ϵ, v_ϵ.

Ainsi ces nouvelles variables vérifient les systèmes d'équations suivant :

$$\begin{aligned}
\partial_t n(a,t) &= -\partial_a n(a,t) - \mu(a)n(a,t),\\
\partial_t w_\epsilon(a,t) &= -\partial_a w_\epsilon(a,t) - \mu(a)w_\epsilon(a,t) + \Lambda(a, l_\epsilon(\cdot,t), x_\epsilon(\cdot,t), z_\epsilon(\cdot,t))\\
&\quad \times (n(a,t) - w_\epsilon(a,t) - v_\epsilon(a,t)),\\
\partial_t z_\epsilon(a,t) &= -\partial_a z_\epsilon(a,t) - \mu(a)z_\epsilon(a,t) - \frac{\gamma_1}{\epsilon} z_\epsilon(a,t)\\
&\quad + \frac{\sigma}{\sigma-\gamma_1}\Lambda(a, l_\epsilon(\cdot,t), x_\epsilon(\cdot,t), z_\epsilon(\cdot,t))(n(a,t) - w_\epsilon(a,t) - v_\epsilon(a,t)),\\
\partial_t x_\epsilon(a,t) &= -\partial_a x_\epsilon(a,t) - \mu(a)x_\epsilon(a,t) - \frac{\gamma_2(a)}{\epsilon} x_\epsilon(a,t) - \frac{\gamma_1}{\epsilon}\cdot\frac{\sigma q(a)}{\sigma-\gamma_1} l_\epsilon(a,t)\\
&\quad + \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\cdot\frac{\gamma_2'(a)}{\gamma_1-\gamma_2(a)} z_\epsilon(a,t) + \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\\
&\quad \times \Lambda(a, l_\epsilon(\cdot,t), x_\epsilon(\cdot,t), z_\epsilon(\cdot,t))(n(a,t) - w_\epsilon(a,t) - v_\epsilon(a,t)),\\
\partial_t l_\epsilon(a,t) &= -\partial_a l_\epsilon(a,t) - \mu(a)l_\epsilon(a,t) - \frac{\sigma}{\epsilon} l_\epsilon(a,t)\\
&\quad + \Lambda(a, l_\epsilon(\cdot,t), x_\epsilon(\cdot,t), z_\epsilon(\cdot,t))(n(a,t) - w_\epsilon(a,t) - v_\epsilon(a,t)),\\
\partial_t v_\epsilon(a,t) &= -\partial_a v_\epsilon(a,t) - \mu(a)v_\epsilon(a,t) - \psi v_\epsilon(a,t) - p(a,t)v_\epsilon(a,t)\\
&\quad + p(a,t)(n(a,t) - w_\epsilon(a,t)),
\end{aligned} \tag{2.24}$$

les conditions aux bords

$$\begin{aligned}
n(0,t) &= \int_0^{a_+} \beta(a)n(a,t)\,da,\\
w_\epsilon(0,t) &= x_\epsilon(0,t) = z_\epsilon(0,t) = l_\epsilon(0,t) = 0,\\
v_\epsilon(0,t) &= (1-\omega)n(0,t),
\end{aligned} \tag{2.25}$$

et les conditions initiales

$$\begin{aligned}
n(a,0) &= \overset{\circ}{n}(a) = \overset{\circ}{s}(a) + \overset{\circ}{l}(a) + \overset{\circ}{i}(a) + \overset{\circ}{c}(a) + \overset{\circ}{r}(a) + \overset{\circ}{v}(a),\\
w_\epsilon(a,0) &= \overset{\circ}{w}(a) = \overset{\circ}{l}(a) + \overset{\circ}{i}(a) + \overset{\circ}{c}(a) + \overset{\circ}{r}(a),\\
z_\epsilon(a,0) &= \overset{\circ}{z}(a) = \frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \overset{\circ}{i}(a),\\
x_\epsilon(a,0) &= \overset{\circ}{x}(a) = \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\cdot\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\overset{\circ}{i}(a) + \overset{\circ}{c}(a),\\
l_\epsilon(a,0) &= \overset{\circ}{l}(a), \quad v_\epsilon(a,0) = \overset{\circ}{v}(a),
\end{aligned} \tag{2.26}$$

où la force d'infection

$$\begin{aligned}
\Lambda(a, l_\epsilon(\cdot,t), x_\epsilon(\cdot,t), z_\epsilon(\cdot,t)) = k(a)\int_0^{a_+}\Bigg[&-\frac{\sigma}{\sigma-\gamma_1}h_1(a)l_\epsilon(a,t) + h_2(a)x_\epsilon(a,t)\\
&+\left(h_1(a) - \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}h_2(a)\right)z_\epsilon(a,t)\Bigg]\,da.
\end{aligned}$$

Cependant, on note que le $(n, w_\epsilon, v_\epsilon) \in V$ et le triplet $(z_\epsilon, x_\epsilon, l_\epsilon) \in W$. Nous voyons que la population totale n est séparée du système, il est inutile de l'approximer et peut être traitée comme une fonction connue. Par conséquent, nous allons nous focaliser sur le quintuplet $(w_\epsilon, z_\epsilon, x_\epsilon, l_\epsilon, v_\epsilon)$.

Soit $(\overline{w}, \overline{z}, \overline{x}, \bar{l}, \overline{v})$ l'approximation globale de $(w_\epsilon, z_\epsilon, x_\epsilon, l_\epsilon, v_\epsilon)$. Suivant la procédure d'approximation globale de Chapman-Enskog [11, 13], nous développons uniquement la partie cinétique de l'approximation globale telle que :

$$\begin{aligned}
\overline{z} &= \overline{z}_0 + \epsilon\overline{z}_1 + \cdots,\\
\overline{x} &= \overline{x}_0 + \epsilon\overline{x}_1 + \cdots,\\
\bar{l} &= \bar{l}_0 + \epsilon\bar{l}_1 + \cdots.
\end{aligned} \tag{2.27}$$

En insérant (2.27) dans les quatre dernières équations de (2.24), nous obtenons

$$\begin{aligned}
\partial_t\overline{z}_0+\epsilon\partial_t\overline{z}_1 &= -\partial_a\overline{z}_0-\epsilon\partial_a\overline{z}_1-\mu\overline{z}_0-\epsilon\mu\overline{z}_1-\frac{\gamma_1}{\epsilon}\overline{z}_0-\gamma_1\overline{z}_1+\frac{\sigma}{\sigma-\gamma_1}\\
&\quad\times\Big(\Lambda(\bar{l}_0,\overline{x}_0,\overline{z}_0)+\epsilon\Lambda(\bar{l}_1,\overline{x}_1,\overline{z}_1)\Big)(n-\overline{w}-\overline{v})+O(\epsilon^2),\\
\partial_t\overline{x}_0+\epsilon\partial_t\overline{x}_1 &= -\partial_a\overline{x}_0-\epsilon\partial_a\overline{x}_1-\mu\overline{x}_0-\epsilon\mu\overline{x}_1-\frac{\gamma_2}{\epsilon}\overline{x}_0-\gamma_2\overline{x}_1\\
&\quad+\frac{1}{\epsilon}\cdot\frac{q\gamma_1^2}{\gamma_1-\gamma_2}\overline{z}_0+\frac{q\gamma_1^2}{\gamma_1-\gamma_2}\overline{z}_1-\frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{q\gamma_2^{'}}{\gamma_1-\gamma_2}\overline{z}_0\\
&\quad-\epsilon\frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{q\gamma_2^{'}}{\gamma_1-\gamma_2}\overline{z}_1-\frac{q}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma-\gamma_1}\bar{l}_0-\frac{\sigma q\gamma_1}{\sigma-\gamma_1}\bar{l}_1\\
&\quad+\frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\sigma}{\sigma-\gamma_1}\Big(\Lambda(\bar{l}_0,\overline{x}_0,\overline{z}_0)+\epsilon\Lambda(\bar{l}_1,\overline{x}_1,\overline{z}_1)\Big)\\
&\quad\times(n-\overline{w}-\overline{v})+O(\epsilon^2),\\
\partial_t\bar{l}_0+\epsilon\partial_t\bar{l}_1 &= -\partial_a\bar{l}_0-\epsilon\partial_a\bar{l}_1-\mu\bar{l}_0-\epsilon\mu\bar{l}_1+\Big(\Lambda(\bar{l}_0,\overline{x}_0,\overline{z}_0)+\epsilon\Lambda(\bar{l}_1,\overline{x}_1,\overline{z}_1)\Big)\\
&\quad\times(n-\overline{w}-\overline{v})-\frac{\sigma}{\epsilon}\bar{l}_0-\sigma\bar{l}_1+O(\epsilon^2).
\end{aligned}$$

En comparant les coefficients ayant les mêmes puissances de ϵ et en utilisant le fait que $\Lambda(0,0,0)=0$, nous obtenons $\overline{z}_0=\overline{x}_0=\bar{l}_0=0$, $\overline{z}_1=\overline{x}_1=\bar{l}_1=0$.

On aboutit enfin à l'approximation globale

$$(n,\overline{w},\overline{z},\overline{x},\bar{l},\overline{v})=(n,\overline{w},0,0,0,\overline{v}).$$

En substituant $l_\epsilon\approx\bar{l}=0$, $x_\epsilon\approx\overline{x}=0$, $z_\epsilon\approx\overline{z}=0$ dans la deuxième équation de (2.24), nous arrivons au système (2.18). De plus, la substitution de $\overline{w}$ dans la dernière équation de (2.24), conduit à (2.19).

Estimation de l'erreur d'approximation globale

L'erreur commise en approximant la solution $(n, w_\epsilon, z_\epsilon, x_\epsilon, l_\epsilon, v_\epsilon)$ par $(n, \overline{w}, 0, 0, 0, \overline{v})$ est notée par

$$\begin{aligned}\overline{\mathbf{E}} = (\overline{e}_w, \overline{e}_z, \overline{e}_x, \overline{e}_l, \overline{e}_v) &= (w_\epsilon - \overline{w}, z_\epsilon - \overline{z}, x_\epsilon - \overline{x}, l_\epsilon - \overline{l}, v_\epsilon - \overline{v}) \\ &= (w_\epsilon - \overline{w}, z_\epsilon, x_\epsilon, l_\epsilon, v_\epsilon - \overline{v}),\end{aligned} \tag{2.28}$$

satisfaisant au système d'équation

$$\begin{aligned}
\partial_t \overline{e}_w &= -\partial_a \overline{e}_w - \mu \overline{e}_w - \Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(\overline{e}_w + \overline{e}_v\right) + \Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(n - \overline{w} - \overline{v}\right), \\
\partial_t \overline{e}_z &= -\partial_a \overline{e}_z - \mu \overline{e}_z - \frac{\gamma_1}{\epsilon}\overline{e}_z - \frac{\sigma}{\sigma - \gamma_1}\Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(\overline{e}_w + \overline{e}_v\right) \\
&\quad + \frac{\sigma}{\sigma - \gamma_1}\Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(n - \overline{w} - \overline{v}\right), \\
\partial_t \overline{e}_x &= -\partial_a \overline{e}_x - \mu \overline{e}_x - \frac{\gamma_2}{\epsilon}\overline{e}_x - \frac{1}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma - \gamma_1}\overline{e}_l + \frac{q\gamma_1}{\gamma_1 - \gamma_2}\cdot\frac{\gamma_2^{'}}{\gamma_1 - \gamma_2}\overline{e}_z \\
&\quad - \frac{\sigma}{\sigma - \gamma_1}\cdot\frac{q\gamma_1}{\gamma_1 - \gamma_2}\Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(\overline{e}_w + \overline{e}_v\right) \\
&\quad + \frac{\sigma}{\sigma - \gamma_1}\cdot\frac{q\gamma_1}{\gamma_1 - \gamma_2}\Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(n - \overline{w} - \overline{v}\right), \\
\partial_t \overline{e}_l &= -\partial_a \overline{e}_l - \mu \overline{e}_l - \frac{\sigma}{\epsilon}\overline{e}_l - \Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(\overline{e}_w + \overline{e}_v\right) \\
&\quad + \Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(n - \overline{w} - \overline{v}\right), \\
\partial_t \overline{e}_v &= -\partial_a \overline{e}_v - \mu \overline{e}_v - (p + \psi)\overline{e}_v - p\overline{e}_w,
\end{aligned} \tag{2.29}$$

avec les conditions aux bords

$$\overline{e}_w(0,t) = \overline{e}_z(0,t) = \overline{e}_x(0,t) = \overline{e}_l(0,t) = \overline{e}_v(0,t) = 0, \tag{2.30}$$

et les conditions initiales

$$\begin{aligned}
&\overline{e}_w(a,0) = 0, \quad \overline{e}_z(a,0) = \frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \overset{\circ}{i}(a),\\
&\overline{e}_x(a,0) = \frac{q\gamma_1}{\gamma_1-\gamma_2(a)}\cdot\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \frac{q\gamma_1}{\gamma_1-\gamma_2(a)}\overset{\circ}{i}(a) + \overset{\circ}{c}(a),\\
&\overline{e}_l(a,0) = \overset{\circ}{l}(a), \quad \overline{e}_v(a,0) = 0.
\end{aligned} \tag{2.31}$$

Nous remarquons que la condition initiale est d'ordre 1, donc l'erreur d'approximation globale ne peut pas être d'ordre ϵ. Pour remédier à cette situation, nous devons introduire des corrections de couche qui prendront en charge les phénomènes transitoires se produisant près de $t = 0$ et $a = 0$ appelés respectivement correction aux valeurs initiales et correction valeurs aux bords.

Correction aux valeurs initiales

Nous effectuons la correction aux valeurs initiales en faisant exploser le temps selon $\tau = t/\epsilon$ et en recherchant l'approximation

$$(w_\epsilon(t), z_\epsilon(t), x_\epsilon(t), l_\epsilon(t), v_\epsilon(t)) \approx (\overline{w}(t), \widetilde{z}(\tau), \widetilde{x}(\tau), \widetilde{l}(\tau), \overline{v}(t)),$$

où nous prévoyons qu'il est inutile d'introduire la correction des valeurs initiales pour w_ϵ et v_ϵ comme $\overline{w}$ et $\overline{v}$ remplissent la condition initiale exacte. Comme précédemment, nous insérons le développement formelle

$$\begin{aligned}
\widetilde{z} &= \widetilde{z}_0 + \epsilon\widetilde{z}_1 + \cdots,\\
\widetilde{x} &= \widetilde{x}_0 + \epsilon\widetilde{x}_1 + \cdots,\\
\widetilde{l} &= \widetilde{l}_0 + \epsilon\widetilde{l}_1 + \cdots
\end{aligned}$$

dans les troisième, quatrième et cinquième équations de (2.24) et en prenant en compte le changement de variable $\partial_t = \epsilon^{-1}\partial_\tau$.

En comparant les coefficients ayant les mêmes puissances de ϵ, les équations pour

les termes ϵ^{-1} sont

$$\partial_\tau \widetilde{z}_0 = -\gamma_1 \widetilde{z}_0,\ d\partial_\tau \widetilde{l}_0 = -\sigma \widetilde{l}_0,$$
$$\partial_\tau \widetilde{x}_0 = -\gamma_2(a)\widetilde{x}_0 - \frac{\sigma\gamma_1}{\sigma-\gamma_1}\widetilde{l}_0$$

sous les conditions initiales

$$\widetilde{z}_0(0) = \frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l} + \overset{\circ}{i}, \quad \widetilde{l}_0(0) = \overset{\circ}{l},$$
$$\widetilde{x}_0(0) = \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)} \cdot \frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l} + \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\overset{\circ}{i} + \overset{\circ}{c},$$

et sous les conditions aux bords

$$\begin{aligned}
\widetilde{z}_0(a,t/\epsilon) &= \left(\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \overset{\circ}{i}(a)\right) e^{-\frac{\gamma_1}{\epsilon}t}, \quad \widetilde{l}_0(a,t/\epsilon) = \overset{\circ}{l}(a)e^{-\frac{\sigma}{\epsilon}t}, \\
\widetilde{x}_0(a,t/\epsilon) &= \left(\frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)} \cdot \frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\overset{\circ}{i}(a) + \overset{\circ}{c}(a)\right) e^{-\frac{\gamma_2(a)}{\epsilon}t} \\
&\quad + \frac{\sigma}{\sigma-\gamma_1} \cdot \frac{\overset{\circ}{l}(a)}{\sigma-\gamma_2(a)}\left(e^{-\frac{\sigma}{\epsilon}t} - e^{-\frac{\gamma_2(a)}{\epsilon}t}\right).
\end{aligned} \tag{2.32}$$

La nouvelle estimation de l'erreur d'approximation est donnée par

$$\begin{aligned}
\widetilde{\mathbf{E}} = (\widetilde{e}_w, \widetilde{e}_z, \widetilde{e}_x, \widetilde{e}_l, \widetilde{e}_v) &= (w_\epsilon - \overline{w}, z_\epsilon - \widetilde{z}_0, x_\epsilon - \widetilde{x}_0, l_\epsilon - \widetilde{l}_0, v_\epsilon - \overline{v}) \\
&= (\overline{e}_w, \overline{e}_z - \widetilde{z}_0, \overline{e}_x - \widetilde{x}_0, \overline{e}_l - \widetilde{l}_0, \overline{e}_v).
\end{aligned} \tag{2.33}$$

On note que $\overset{\circ}{l}, \overset{\circ}{i}, \overset{\circ}{c}, \overset{\circ}{v} \in \mathrm{W}^{1,1}([0,a_+])$ et $\mu\overset{\circ}{l}, \mu\overset{\circ}{i}, \mu\overset{\circ}{c}, \mu\overset{\circ}{v} \in \mathrm{L}^1([0,a_+])$.

L'équation d'erreurs pour $\widetilde{E}$ s'obtient à partir de (2.29), (2.30) et (2.31) en exprimant

$\overline{e}_w$, $\overline{e}_z$, $\overline{e}_x$, $\overline{e}_l$ et $\overline{e}_v$ en terme de $\widetilde{e}_w$, $\widetilde{e}_z$, $\widetilde{e}_x$, $\widetilde{e}_l$ et $\widetilde{e}_v$, selon (2.33). Ainsi nous obtenons

$$\begin{aligned}
\partial_t\widetilde{e}_w &= -\partial_a\widetilde{e}_w - \mu\widetilde{e}_w - \Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)\,(\widetilde{e}_w+\widetilde{e}_v) - \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(\widetilde{e}_w+\widetilde{e}_v)\\
&\quad + \Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)\,(n-\overline{w}-\overline{v}) - \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(n-\overline{w}-\overline{v}),\\
\partial_t\widetilde{e}_z &= -\partial_a\widetilde{e}_z - \mu\widetilde{e}_z - \frac{\gamma_1}{\epsilon}\widetilde{e}_z - \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)\,(\widetilde{e}_w+\widetilde{e}_v)\\
&\quad - \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(\widetilde{e}_w+\widetilde{e}_v) + \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)\,(n-\overline{w}-\overline{v})\\
&\quad - \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(n-\overline{w}-\overline{v}) - \partial_a\widetilde{z}_0 - \mu\widetilde{z}_0,\\
\partial_t\widetilde{e}_x &= -\partial_a\widetilde{e}_x - \mu\widetilde{e}_x - \frac{\gamma_2}{\epsilon}\widetilde{e}_x - \frac{1}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma-\gamma_1}\widetilde{e}_l + \frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\gamma_2'}{\gamma_1-\gamma_2}\widetilde{e}_z\\
&\quad - \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}\Big(\Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)+\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\Big)\,(\widetilde{e}_w+\widetilde{e}_v)\\
&\quad + \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}\Big(\Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)-\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\Big)\,(n-\overline{w}-\overline{v})\\
&\quad + \frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\gamma_2'}{\gamma_1-\gamma_2}\widetilde{z}_0 - \partial_a\widetilde{x}_0 - \mu\widetilde{x}_0,\\
\partial_t\widetilde{e}_l &= -\partial_a\widetilde{e}_l - \mu\widetilde{e}_l - \frac{\sigma}{\epsilon}\widetilde{e}_l - \Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)\,(\widetilde{e}_w+\widetilde{e}_v) - \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(\widetilde{e}_w+\widetilde{e}_v)\\
&\quad + \Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)\,(n-\overline{w}-\overline{v}) - \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(n-\overline{w}-\overline{v}) - \partial_a\widetilde{l}_0 - \mu\widetilde{l}_0,\\
\partial_t\widetilde{e}_v &= -\partial_a\widetilde{e}_v - \mu\widetilde{e}_v - (p+\psi)\widetilde{e}_v - p\widetilde{e}_w,
\end{aligned} \tag{2.34}$$

avec les conditions aux bords

$$\begin{aligned}
&\widetilde{e}_w(0,t) = 0, \quad \widetilde{e}_z(0,t) = -\widetilde{z}_0(0,t/\epsilon), \quad \widetilde{e}_x(0,t) = -\widetilde{x}_0(0,t/\epsilon),\\
&\widetilde{e}_l(0,t) = -\widetilde{l}_0(0,t/\epsilon), \quad \widetilde{e}_v(0,t) = 0,
\end{aligned} \tag{2.35}$$

et les conditions initiales

$$\widetilde{e}_w(a,0) = \widetilde{e}_z(a,0) = \widetilde{e}_x(a,0) = \widetilde{e}_l(a,0) = \widetilde{e}_v(a,0) = 0. \tag{2.36}$$

Nous remarquons qu'aux bords nous avons encore des termes d'ordre inférieur à ϵ sauf $\widetilde{e}_w(0,t) = 0$ et $\widetilde{e}_v(0,t) = 0$. Heureusement, pour éliminer cette contribution aux valeurs initiales, nous introduisons des corrections aux valeurs du bord en

dimensionnant simultanément le temps et l'âge tel que $\tau = t/\epsilon$ et de $\alpha = a/\epsilon$.

Correction aux valeurs du bord

L'approche standard de la correction au bord, [10], ne suffira pas ici malheureusement car les équations des corrections aux bords n'intègrent pas la multiplication par μ. Ainsi, la correction au bord classique n'appartiendra pas à $D(\mathcal{M})$ et nous ne pourrons pas substituer les termes d'erreur aux équations, comme dans (2.34) - (2.36). Pour remédier au problème, nous définissons le correcteur au bord d'être la solution de

$$\begin{aligned}
\partial_t \breve{z} &= -\partial_a \breve{z} - \mu \breve{z} - \frac{\gamma_1}{\epsilon}\breve{z}, \\
\partial_t \breve{x} &= -\partial_a \breve{x} - \mu \breve{x} - \frac{\gamma_2}{\epsilon}\breve{x} - \frac{\gamma_1}{\epsilon}\cdot\frac{q\sigma}{\sigma-\gamma_1}\breve{l}, \\
\partial_t \breve{l} &= -\partial_a \breve{l} - \mu \breve{l} - \frac{\sigma}{\epsilon}\breve{l},
\end{aligned} \tag{2.37}$$

avec les conditions aux bords

$$\breve{z}(0,t) = -\widetilde{z}_0(0,t/\epsilon), \quad \breve{x}(0,t) = -\widetilde{x}_0(0,t/\epsilon), \quad \breve{l}(0,t) = -\widetilde{l}_0(0,t/\epsilon), \tag{2.38}$$

et les conditions initiales

$$\breve{z}(a,0) = c_\epsilon e^{-\frac{a}{\epsilon}}, \quad \breve{x}(a,0) = d_\epsilon e^{-\frac{a}{\epsilon}}, \quad \breve{l}(a,0) = h_\epsilon e^{-\frac{a}{\epsilon}}, \tag{2.39}$$

où c_ϵ, d_ϵ et h_ϵ sont des constantes obtenues à partir de l'égalité des conditions au bord et initiale $(a,t) = (0,0)$, telle que la résolution classique du problème avec une condition au bord non homogène soit vérifiée. Par conséquent,

$$\begin{aligned}
c_\epsilon &= -\widetilde{z}_0(0,0) = -\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(0) - \overset{\circ}{i}(0), \quad h_\epsilon = -\overset{\circ}{l}(0), \\
d_\epsilon &= -\widetilde{x}_0(0,0) = -\frac{q(0)\gamma_1}{\gamma_1-\gamma_2(0)}\cdot\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(0) - \frac{q(0)\gamma_1}{\gamma_1-\gamma_2(0)}\overset{\circ}{i}(0) - \overset{\circ}{c}(0)
\end{aligned} \tag{2.40}$$

comme nous avons supposé que la condition initiale pour le problème initial satisfait

$(\overset{\circ}{s}, \overset{\circ}{l}, \overset{\circ}{i}, \overset{\circ}{c}, \overset{\circ}{r}, \overset{\circ}{v}) \in D(\mathcal{A})$.

Les estimations nécessaires de la solution de la correction au bord seront fournies ultérieurement. Ici, en supposant qu'elle soit suffisamment régulière, on considère la nouvelle approximation

$$(w_\epsilon, z_\epsilon, x_\epsilon, l_\epsilon, v_\epsilon) \approx (\overline{w}, \widetilde{z}_0 + \breve{l}, \widetilde{x}_0 + \breve{i}, \widetilde{l}_0 + \breve{c}, \overline{v}).$$

et l'erreur correspondante,

$$\begin{aligned}
\breve{E} &= (\breve{e}_w, \breve{e}_z, \breve{e}_x, \breve{e}_l, \breve{e}_v) \\
&= (w_\epsilon - \overline{w}, z_\epsilon - \widetilde{z}_0 - \breve{z}, x_\epsilon - \widetilde{x}_0 - \breve{x}, l_\epsilon - \widetilde{l}_0 - \breve{l}, v_\epsilon - \overline{v}) \\
&= (\widetilde{e}_w, \widetilde{e}_z - \breve{z}, \widetilde{e}_x - \breve{x}, \widetilde{e}_l - \breve{l}, \widetilde{e}_v),
\end{aligned}$$

vérifie le système

$$
\begin{aligned}
\partial_t \breve{e}_w &= -\partial_a \breve{e}_w - \mu \breve{e}_w - \Lambda(\breve{e}_l, \breve{e}_x, \breve{e}_z)\,(\breve{e}_w + \breve{e}_v) - \Lambda(\breve{l}, \breve{x}, \breve{z})\,(\breve{e}_w + \breve{e}_v) \\
&\quad + \Lambda(\breve{e}_l, \breve{e}_x, \breve{e}_z)\,(n - \overline{w} - \overline{v}) - \Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)\,(\breve{e}_w + \breve{e}_v) \\
&\quad - \Lambda(\breve{l}, \breve{x}, \breve{z})\breve{w} + \Lambda(\breve{l}, \breve{x}, \breve{z})\,(n - \overline{w} - \overline{v}) \\
&\quad + \Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)\,(n - \overline{w} - \overline{v})\,, \\
\partial_t \breve{e}_z &= -\partial_a \breve{e}_z - \mu \breve{e}_z - \frac{\gamma_1}{\epsilon}\breve{e}_z - \frac{\sigma}{\sigma - \gamma_1}\Lambda(\breve{e}_l, \breve{e}_x, \breve{e}_z)\,(\breve{e}_w + \breve{e}_v) \\
&\quad - \frac{\sigma}{\sigma - \gamma_1}\Lambda(\breve{l}, \breve{x}, \breve{z})\,(\breve{e}_w + \breve{e}_v) + \frac{\sigma}{\sigma - \gamma_1}\Lambda(\breve{e}_l, \breve{e}_x, \breve{e}_z)\,(n - \overline{w} - \overline{v}) \\
&\quad - \frac{\sigma}{\sigma - \gamma_1}\Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)\,(\breve{e}_w + \breve{e}_v) + \frac{\sigma}{\sigma - \gamma_1}\Lambda(\breve{l}, \breve{x}, \breve{z})\,(n - \overline{w} - \overline{v}) \\
&\quad + \frac{\sigma}{\sigma - \gamma_1}\Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)\,(n - \overline{w} - \overline{v}) - \partial_a \widetilde{z}_0 - \mu \widetilde{z}_0, \\
\partial_t \breve{e}_x &= -\partial_a \breve{e}_x - \mu \breve{e}_x - \frac{\gamma_2}{\epsilon}\breve{e}_x - \frac{1}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma - \gamma_1}\breve{e}_l + \frac{q\gamma_1}{\gamma_1 - \gamma_2}\cdot\frac{\gamma_2'}{\gamma_1 - \gamma_2}\breve{e}_z \\
&\quad - \frac{\sigma}{\sigma - \gamma_1}\cdot\frac{q\gamma_1}{\gamma_1 - \gamma_2}\Big(\Lambda(\breve{e}_l, \breve{e}_x, \breve{e}_z) + \Lambda(\breve{l}, \breve{x}, \breve{z})\Big)\,(\breve{e}_w + \breve{e}_v) \\
&\quad + \frac{\sigma}{\sigma - \gamma_1}\cdot\frac{q\gamma_1}{\gamma_1 - \gamma_2}\Big(\Lambda(\breve{e}_l, \breve{e}_x, \breve{e}_z)\,(n - \overline{w} - \overline{v}) - \Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)\,(\breve{e}_w + \breve{e}_v)\Big) \\
&\quad + \frac{\sigma}{\sigma - \gamma_1}\cdot\frac{q\gamma_1}{\gamma_1 - \gamma_2}\Big(\Lambda(\breve{l}, \breve{x}, \breve{z}) + \Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)\Big)\,(n - \overline{w} - \overline{v}) \\
&\quad + \frac{q\gamma_1}{\gamma_1 - \gamma_2}\cdot\frac{\gamma_2'}{\gamma_1 - \gamma_2}\,(\breve{z} + \widetilde{z}_0) - \frac{1}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma - \gamma_1}\breve{l} - \partial_a \widetilde{x}_0 - \mu \widetilde{x}_0, \\
\partial_t \breve{e}_l &= -\partial_a \breve{e}_l - \mu \breve{e}_l - \frac{\sigma}{\epsilon}\breve{e}_l - \Lambda(\breve{e}_l, \breve{e}_x, \breve{e}_z)\,(\breve{e}_w + \breve{e}_v) - \Lambda(\breve{l}, \breve{x}, \breve{z})\,(\breve{e}_w + \breve{e}_v) \\
&\quad + \Big(\Lambda(\breve{e}_l, \breve{e}_x, \breve{e}_z) + \Lambda(\breve{l}, \breve{x}, \breve{z})\Big)\,(n - \overline{w} - \overline{v}) - \Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)\,(\breve{e}_w + \breve{e}_v) \\
&\quad + \Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)\,(n - \overline{w} - \overline{v}) - \partial_a \widetilde{l}_0 - \mu \widetilde{l}_0, \\
\partial_t \breve{e}_v &= -\partial_a \breve{e}_v - \mu \breve{e}_v - (p + \psi)\breve{e}_v - p\breve{e}_w,
\end{aligned}
\tag{2.41}
$$

avec les conditions aux bords

$$
\breve{e}_w(0,t) = \breve{e}_z(0,t) = \breve{e}_x(0,t) = \breve{e}_l(0,t) = \breve{e}_v(0,t) = 0, \tag{2.42}
$$

et les conditions initiales

$$\begin{aligned} &\breve{e}_w(a,0) = 0, \quad \breve{e}_z(a,0) = -c_\epsilon e^{-\frac{a}{\epsilon}}, \quad \breve{e}_x(a,0) = -d_\epsilon e^{-\frac{a}{\epsilon}}, \\ &\breve{e}_l(a,0) = -h_\epsilon e^{-\frac{a}{\epsilon}}, \quad \breve{e}_v(a,0) = 0. \end{aligned} \tag{2.43}$$

Nous remarquons enfin qu'aux bords tous les termes non nuls sont d'ordre ϵ. Pour achever ce chapitre nous donnons dans la suite une série de résultats théoriques pour justifier cette convergence. Le lecteur pourra se référer à ([36]) pour les preuves de ces résultats.

Lemme 2.2 *Soit $M_\theta = \mu + \theta/\epsilon$, où θ est une constante. Si η est la solution de*

$$\begin{aligned} \partial_t \eta &= -\partial_a \eta - M_\theta \eta, \\ \eta(0,t) &= -\delta_1 e^{-\frac{\theta}{\epsilon}t}, \\ \eta(a,0) &= -\delta_2 e^{-\frac{a}{\epsilon}}, \end{aligned}$$

alors il existe une constante non nulle C_η telle que

$$\|\eta(t)\| \leq \epsilon C_\eta e^{-\frac{\theta}{2\epsilon}t}. \tag{2.44}$$

A partir du résultat précédent, nous pouvons majorer $\breve{z}$ et $\breve{l}$, solution de (2.37)-(2.39), où $\delta_1 = \delta_2$. Nous avons le résultat suivant :

Proposition 2.1 *Soient $\breve{z}$ et $\breve{l}$ les solutions de* (2.37)-(2.39)*. Alors ils existent des constantes K_z et dépendant des coefficients des valeurs initiales $\overset{\circ}{i}$ et $\overset{\circ}{l}$ et de la norme de* $\mathrm{W}^{1,1}(\mathbb{R}_+)$ *, telle que pour tout $t \in \mathbb{R}_+$,*

$$\|\breve{z}(t)\| \leq \epsilon K_z e^{-\frac{\gamma_1}{2\epsilon}t}, \tag{2.45}$$

$$\|\breve{l}(t)\| \leq \epsilon K_l e^{-\frac{\sigma}{2\epsilon}t}, . \tag{2.46}$$

En somme l'étude de l'approximation asymptotique du système 2.11 aboutit au résultat fondamental suivant :

Théorème 2.3 *Considérant que les coefficients du problème (2.11) vérifient les conditions* $\mathbf{A1}--\mathbf{A4}$ *et (2.17) et la condition initiale* $(\overset{\circ}{s}, \overset{\circ}{l}, \overset{\circ}{i}, \overset{\circ}{c}, \overset{\circ}{r}, \overset{\circ}{v})$ *est telle que* $(s_\epsilon, l_\epsilon, i_\epsilon, c_\epsilon, r_\epsilon, v_\epsilon)$ *est une solution classique de* (2.11)*, alors, il existe des constantes* $C_1, C_2, C_3, C_4, C_5, C_6$*, dépendants seulement des coefficients du problème et la norme de la condition initiale* $D(\mathcal{S}) \cap D(\mathcal{M})$*, telles que pour tout* $\epsilon > 0$ *suffisamment petit*

$$\|s_\epsilon(t) - (n(t) - \overline{w}(t) - \overline{v}(t))\| \leq \epsilon C_1, \tag{2.47}$$

$$\|l_\epsilon(t) - \overset{\circ}{l} e^{-\frac{\sigma}{\epsilon}t}\| \leq \epsilon C_2, \tag{2.48}$$

$$\left\| i_\epsilon(t) - \left(\overset{\circ}{i} e^{-\frac{\gamma_1}{\epsilon}t} + \frac{\sigma \overset{\circ}{l}}{\sigma - \gamma_1} \left(e^{-\frac{\gamma_1}{\epsilon}t} - e^{-\frac{\sigma}{\epsilon}t} \right) \right) \right\| \leq \epsilon C_3, \tag{2.49}$$

$$\begin{aligned} \left\| c_\epsilon(t) - \left(\overset{\circ}{c} e^{-\frac{\gamma_2}{\epsilon}t} - \frac{q\gamma_1^2}{(\gamma_1 - \gamma_2)^2} \left(\frac{\sigma \overset{\circ}{l}}{\sigma - \gamma_1} + \overset{\circ}{i} \right) \left(e^{-\frac{\gamma_2}{\epsilon}t} - e^{-\frac{\gamma_1}{\epsilon}t} \right) \right. \right. \\ \left. \left. - \frac{\sigma \overset{\circ}{l}}{(\sigma - \gamma_1)(\sigma - \gamma_2)} \left(e^{-\frac{\gamma_2}{\epsilon}t} - e^{-\frac{\sigma}{\epsilon}t} \right) \right) \right\| \leq \epsilon C_4, \end{aligned} \tag{2.50}$$

$$\begin{aligned} \left\| r_\epsilon(t) - \left(\overline{w} - \frac{\gamma_1 \overset{\circ}{l}}{\sigma - \gamma_1} e^{-\frac{\sigma}{\epsilon}t} - \frac{(1-q)\gamma_1 - \gamma_2}{\gamma_1 - \gamma_2} \left(\frac{\sigma \overset{\circ}{l}}{\sigma - \gamma_1} + \overset{\circ}{i} \right) e^{-\frac{\gamma_1}{\epsilon}t} \right. \right. \\ - \frac{\sigma \overset{\circ}{l}}{(\sigma - \gamma_1)(\gamma_1 - \gamma_2)} \left(e^{-\frac{\gamma_2}{\epsilon}t} - e^{-\frac{\sigma}{\epsilon}t} \right) \\ - \left(\frac{q\sigma\gamma_1 \overset{\circ}{l}}{(\sigma - \gamma_1)(\gamma_1 - \gamma_2)} + \frac{q\gamma_1 \overset{\circ}{i}}{\gamma_1 - \gamma_2} + \overset{\circ}{c} \right) e^{-\frac{\gamma_1}{\epsilon}t} \\ \left. \left. - \frac{q\gamma_1^2}{(\gamma_1 - \gamma_2)^2} \left(\frac{\sigma \overset{\circ}{l}}{\sigma - \gamma_1} + \overset{\circ}{i} \right) \left(e^{-\frac{\gamma_2}{\epsilon}t} - e^{-\frac{\gamma_1}{\epsilon}t} \right) \right) \right\| \leq \epsilon C_5, \end{aligned} \tag{2.51}$$

$$\|v_\epsilon(t) - \overline{v}(t)\| \leq \epsilon C_6. \tag{2.52}$$

Ce théorème fondamental achève l'estimation de l'erreur d'approximation à ϵ près en utilisant les corrections aux valeurs initiales et aux bords. La preuve de ce théorème est à retrouver dans ([36]).

2.3 Conclusion

Au terme de ce chapitre nous avons dans un premier temps développé un schéma numérique qui a donné une bonne approximation du système d'équation dépendant de l'âge de la population. L'analyse a montré que le taux de convergence est de l'ordre de $\mathcal{O}(h^2)$. De plus, l'ordre de convergence est indépendant de l'intervalle d'âge. Si l'âge maximum n'est pas fini, cela est suffisant de choisir un L de telle sorte que $\operatorname{supp} u^0 \subset [0, hL]$.

Dans un second temps, nous avons présenté un nouveau modèle mathématique de la dynamique de transmission de l'hépatite B dans la population. ce modèle diffère des précédents par sa force d'infection. Ensuite nous avons approximé la solution à travers une analyse asymptotique de type Chapman-Enskog. L'erreur d'approximation obtenue est de l'ordre de ϵ. La suite de ce travail consistera à trouver une solution numérique des erreurs d'approximation commises dans le processus de l'analyse asymptotique. La méthode de Heun proposée dans ce chapitre serait une voie ouverte à cet effet.

Chapitre 3

Interfaces graphiques de résolution des problèmes en environnement et en épidémiologie

Sommaire

Dans ce chapitre, nous présentons un outil de simulation numérique de modèles mathématiques. Nous ferons une brève présentation des interfaces graphiques regroupées en deux sous groupes : l'interface graphique pour les problèmes environnementaux et celle dédiée pour les problèmes épidémiologiques. Des exemples de démonstrations sont donnés dans chaque catégorie.

3.1 Présentation de l'interface graphique

L'outil numérique que nous présentons dans ce travail est une plateforme de résolution de problèmes en environnement et en épidémiologie. Il a été conçu à partir du logiciel Matlab grâce à son application intégrée *GUIDE (Graphical User Interface for the Development Environment)*. Le GUI (Graphical User Interface) résume la plupart des programmes développés dans la résolution des problèmes abordés dans les thématiques citées plus haut. Ainsi, La plateforme comporte au total huit (8) interfaces graphiques qui inter-réagissent entre elles selon la Figure 3.1. Les flèches entres les interfaces indiquent que l'utilisateur a la possibilité de se déplacer entre elles.

Nous avons un GUI ou interface d'entrée, qui constitue la fenêtre d'accueil à laquelle est rattachée toutes les autres et un GUI de sortie qui constitue la fenêtre de sortie de tous les résultats. Les autres GUI sont organisés selon la thématique du problème. Aussi quelque soit la nature du problème, l'utilisateur a la possibilité de chercher une solution approchée du système d'équations (Problème direct) ou d'identifier de manière complète un système d'équations issues d'une modélisation(Problème inverse). Nous procéderons dans la suite à la présentation des différentes interfaces.

Interface(1) : Simulation Principale (Figure 3.2)

Cette interface représente le portail d'accueil, qui permet à l'utilisateur de choisir la thématique du problème à résoudre. Elle comporte les touches suivantes :

- **Environnement** : pour le traitement des modèles environnementaux ;
- **Épidémiologie** : pour le traitement des modèles épidémiologiques et
- **FIN** : pour arrêter le processus.

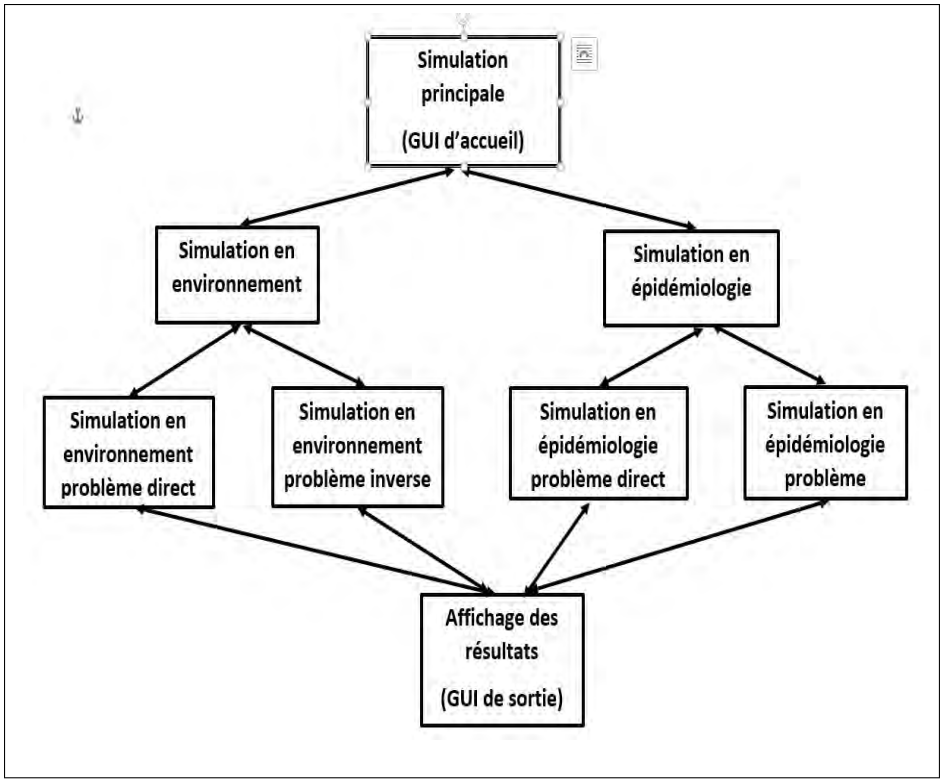


Figure 3.1 – *Organisation et interactions entre les interfaces graphiques.*

Figure 3.2 – *Interface d'accueil.*

3.2 Les interfaces graphiques pour les problèmes environnementaux

Une fois le choix de *Environnemental* fait dans Figure 3.2, nous sommes dirigés vers des interfaces secondaires dédiées à la résolution des modèles issus en environnement et tous les outils numériques servants à leur traitement. Ainsi, nous avons :

Interface(2a) : Simulation en environnement(Figure 3.3)

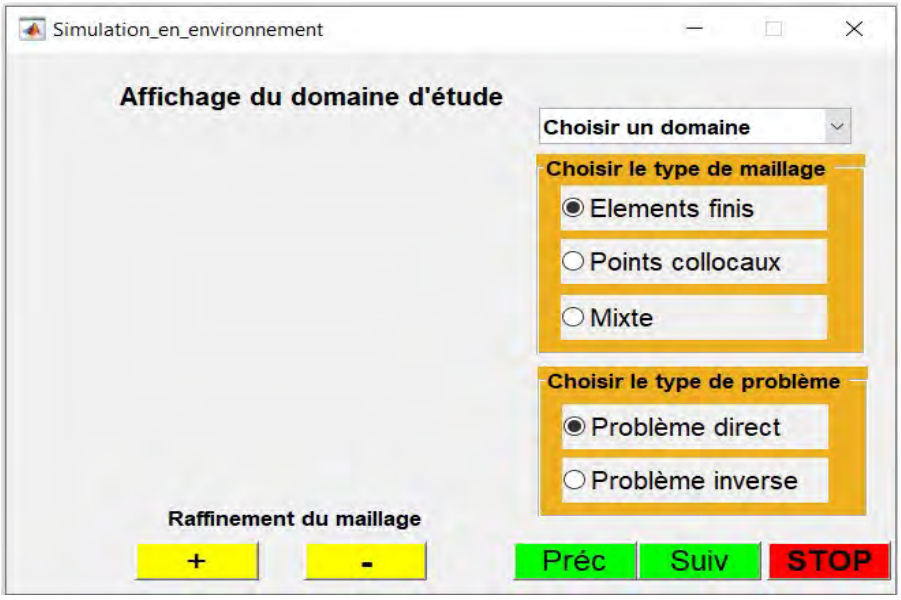


Figure 3.3 – *Interface graphique du choix de maillage et du type de problème.*

Cette fenêtre est la première qui apparaît après la sélection de la touche *Environnemental* dans l'interface Figure 3.2. Elle nous permet de choisir un modèle de domaine (défini par défaut), du type de problème (Problème direct ou inverse), du type de maillage (Eléments finis triangulaires, Points collocaux ou Mixtes). Ensuite, un espace graphique situé à gauche de l'écran nous permet de visualiser le domaine d'étude. Enfin, on a les touches :

+ : pour le raffinement du maillage ;

- : pour le grossissement du maillage ;

Préc. : pour revenir à l'étape précédente (Figure 3.2) ;

Suiv. : pour passer à la prochaine étape (Figures 3.4 ou 3.7) ;

STOP : pour arrêter le processus.

Interface(2a(i)) : Simulation en environnement Pb Direct (Figure 3.4)

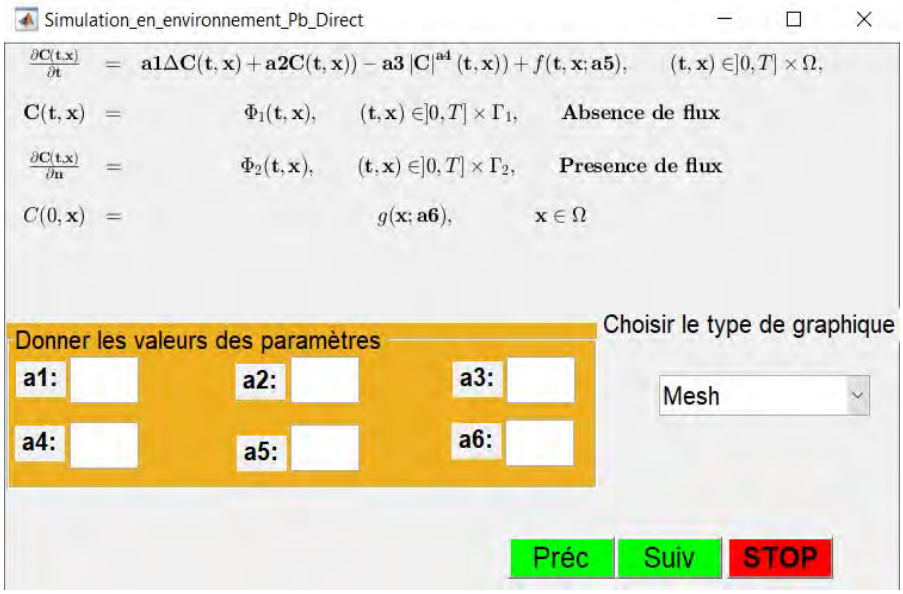


Figure 3.4 – *Interface graphique de résolution des problèmes directs en environnement.*

Cette fenêtre vient à la suite de la Figure 3.3. Ce GUI nous permet d'approcher la solution d'un modèle d'équations paraboliques visibles sur l'écran et comportant des paramètres supposés connus (**ai**, $i = 1, .., 6$). L'opérateur devra entrer les valeurs de ces paramètres, ensuite choisir le type de représentation graphique (*Mesh, Surf, contour ou quiver*) et enfin appuyer sur la touche :

Préc. : pour revenir à l'étape précédente et changer de type de problème ;

Suiv. : pour visualiser la solution approchée du problème ou

STOP : pour arrêter le processus.

Dans Figure 3.5, nous présentons des exemples de maillages à l'aide de l'interface Figure 3.4. Dans cet exemple nous avons choisi deux domaines différents (Domaine 1 et Domaine 2), pris comme type de maillages les éléments finis triangulaires et les éléments finis points collocaux et enfin nous avons procédé au raffinement de chaque type de maillage.

Interface(2a(ii)) : Simulation en environnement Pb Inverse (Figure 3.7)

Cette fenêtre suit la Figure 3.3 lorsque l'utilisateur décide de résoudre un problème inverse. Ce GUI ressemble à Figure 3.4 en ce sens qu'ils présentent tous les deux un modèle d'équations aux dérivées partielles avec des paramètres. Dans le cas présent, les paramètres sont inconnus et l'opérateur devra télécharger des données sur la solution recherchée et faire le choix d'une méthode de résolution pour calculer des valeurs approchées aux dits paramètres. Nous retrouvons également les touches *Prec, Suiv et STOP* définis comme précédemment.

Dans la Figure 3.8, nous donnons des exemples de simulations numériques obtenues en choisissant la méthode des Eléments Finis Points Collocaux. Les données prédéfinis dans cet exemple, ont été celle collectées dans le cas de l'article [46]. Ainsi pour le choix des données *data1* (Figure 3.8(a)), nous obtenons les résultats dans la Figure 3.8(b). Il en est de même pour les figures 3.8(c) et 3.8(d).

3.3 Les interfaces graphiques pour les problèmes épidémiologiques

Une fois le choix de *Epidémiologique* fait dans Figure 3.2, nous nous dirigeons vers des interfaces secondaires dédiées à la résolution des modèles issus dans ce

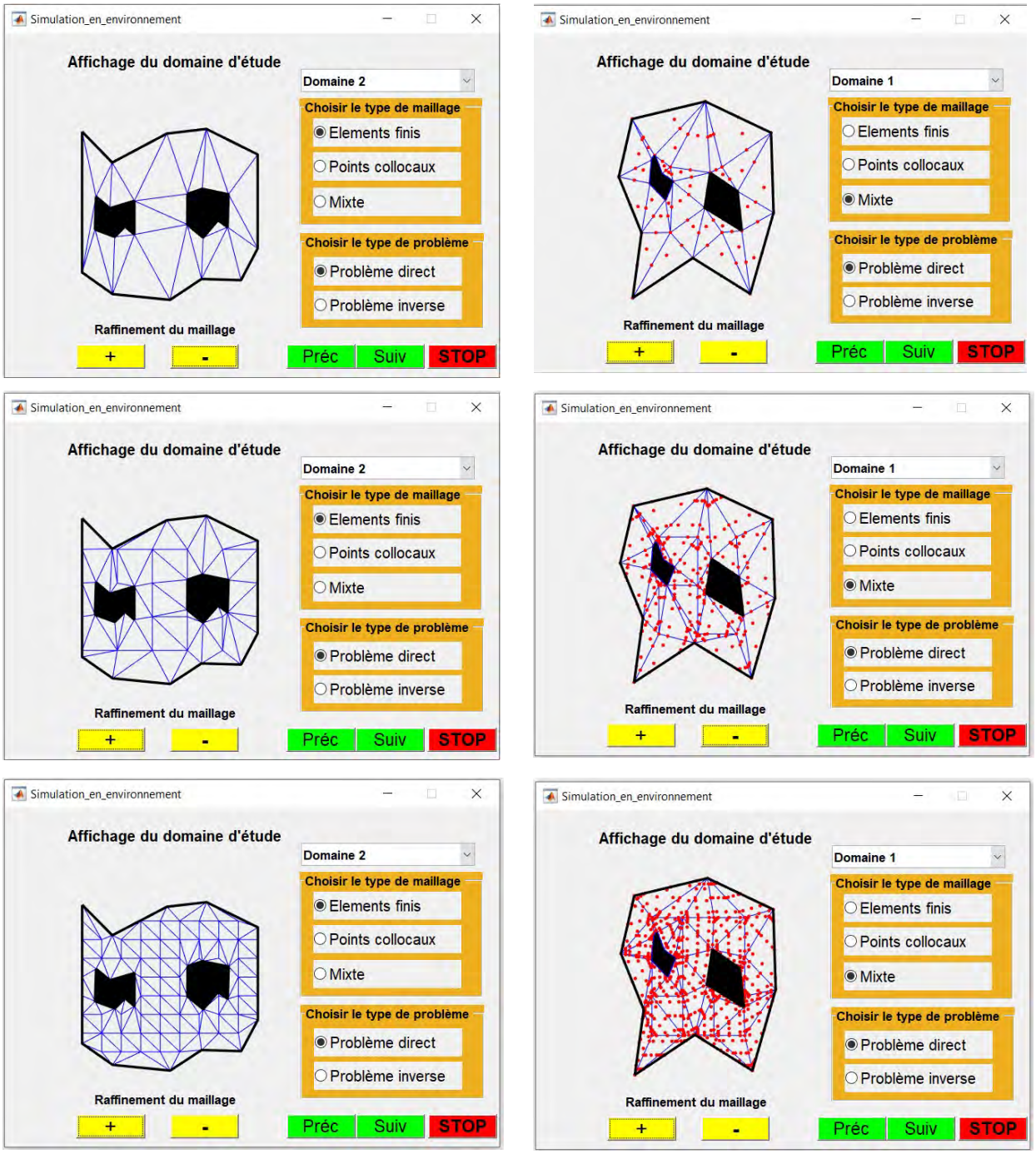


Figure 3.5 – *Exemples de raffinement de maillage du domaine* 1 *par les élements finis (Colonne à gauche) et du domaine* 2 *par les éléments finis collocaux (Colonne à droite).*

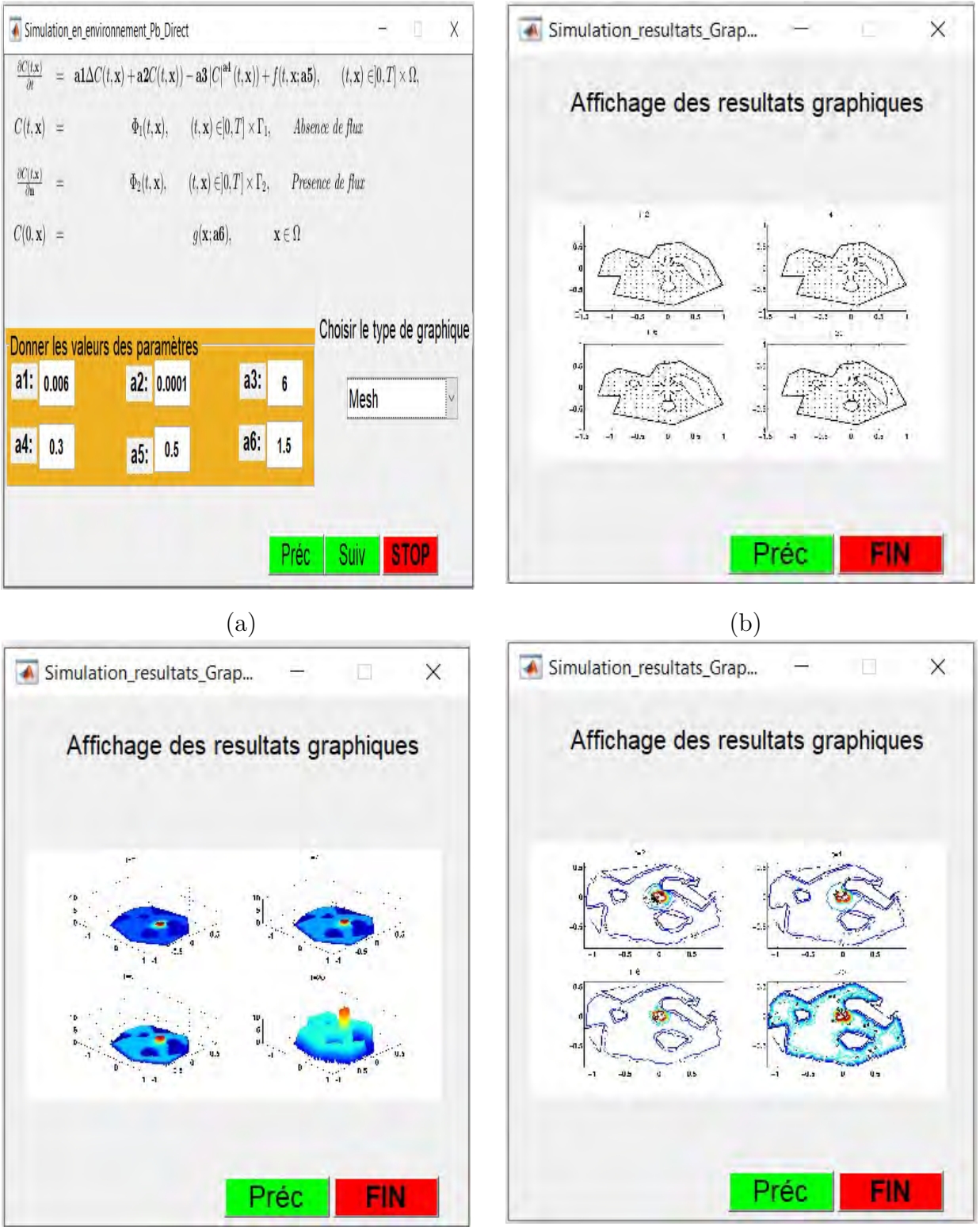


Figure 3.6 – *Resultats de simulation d'un problème direct selon les différents types de graphiques(a) : quiver(b), mesh(c) et contour(d).*

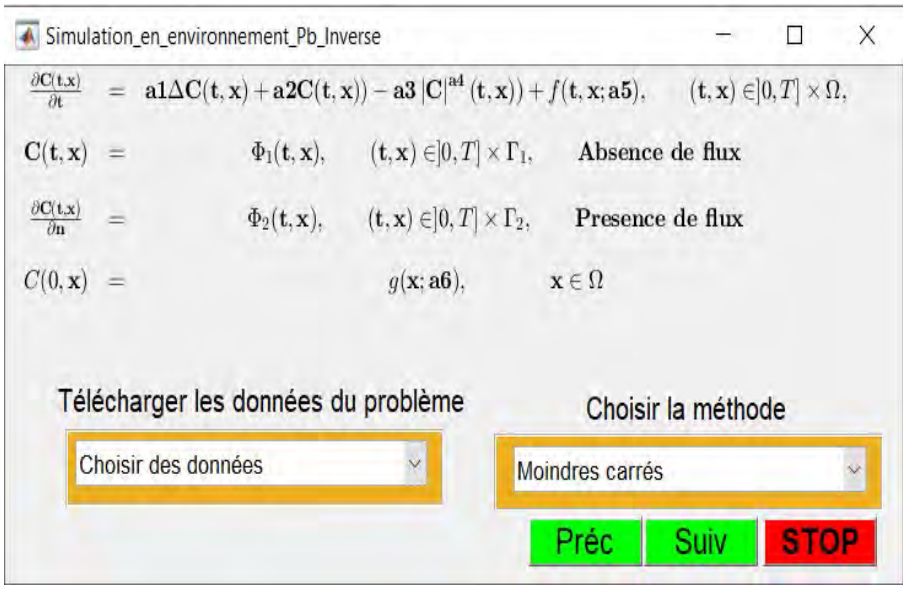


Figure 3.7 – *Interface graphique de résolution de problèmes inverses en environnement.*

domaine et tous les outils numériques servants à leur traitement. Ainsi nous avons :

Interface(2b) : Simulation en Epidemiologie (Figure 3.9)

Dans cette section, nous avons la fenêtre d'accueil (Figure 3.9 (a)) liée à la simulation en épidémiologie. Elle apparaît lorsque nous faisons le choix de la touche *Épidémiologique* dans l'interface Figure 3.2. Elle nous permet de faire le choix du type de problème (Problème direct ou Inverse) et du type de modèle (SIS, SIR,SLICRV). Ainsi pour le choix d'un modèle SIS nous avons l'interface Figure 3.9 (b), pour celui d'un modèle SIR on a Figure 3.9 (c) et enfin pour un modèle SLICRV la plate forme affiche Figure 3.9 (d). Un espace graphique existe pour visualiser le modèle compartimental associé au type de modèle choisi. On a les touches :

Prec. : pour revenir à l'étape précédente (Figure 3.2) ;

Suiv. : pour passer à la prochaine étape(Figures 3.10 ou 3.12) ;

STOP : pour arrêter le processus.

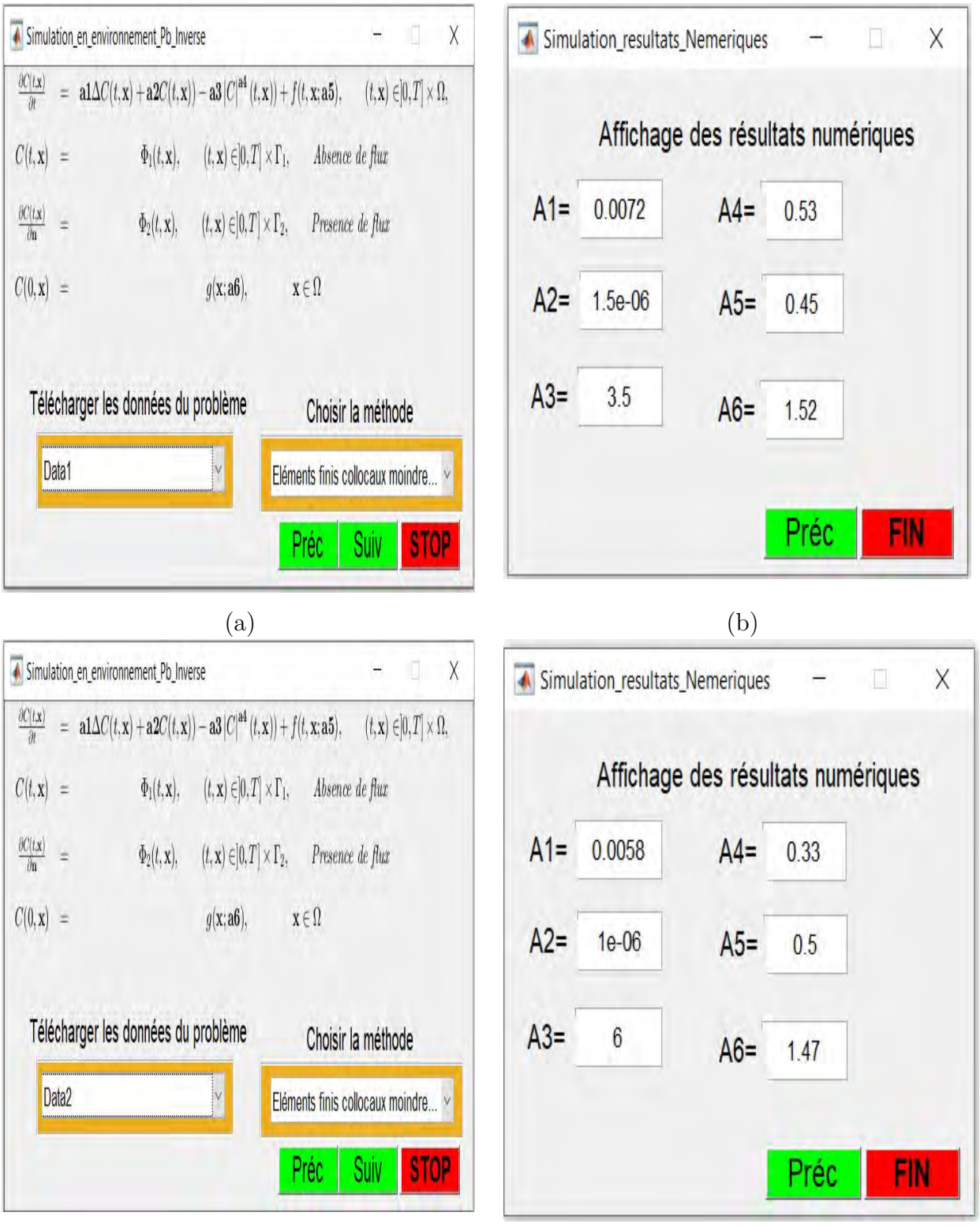


Figure 3.8 – *Résultats de simulation d'un problème inverse : (A gauche)les interfaces indiquant le choix des données et de la méthode, (A droite) les résultats obtenus.*88

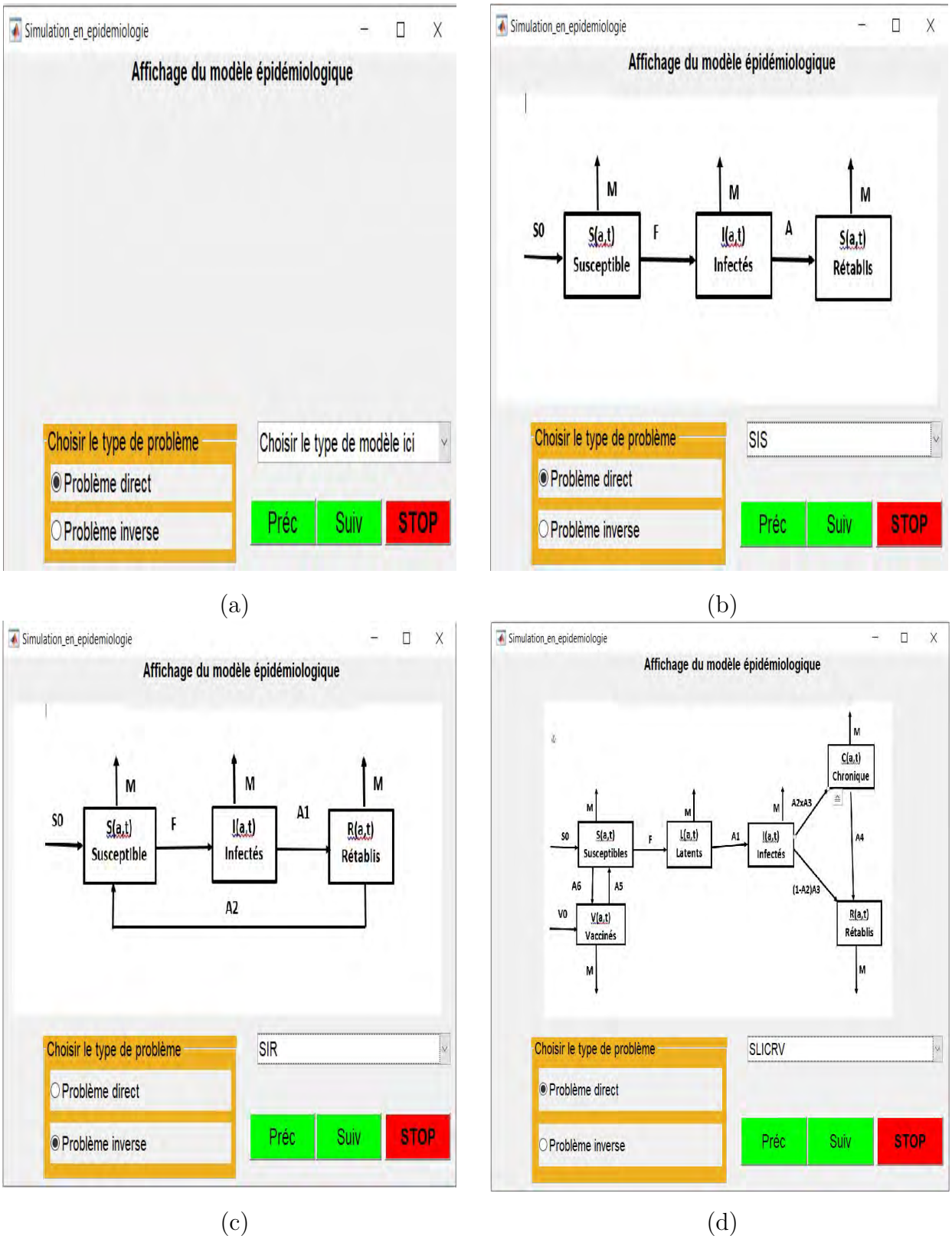


Figure 3.9 – *(a) Interface d'accueil, (b) Interface du modèle SIS, (c) Interface pour le modèle SIR et (d) Interface pour le modèle SLICRV.*

Interface(2b(i)) : Simulation en Epidemiologie Pb Direct (Figure 3.10)

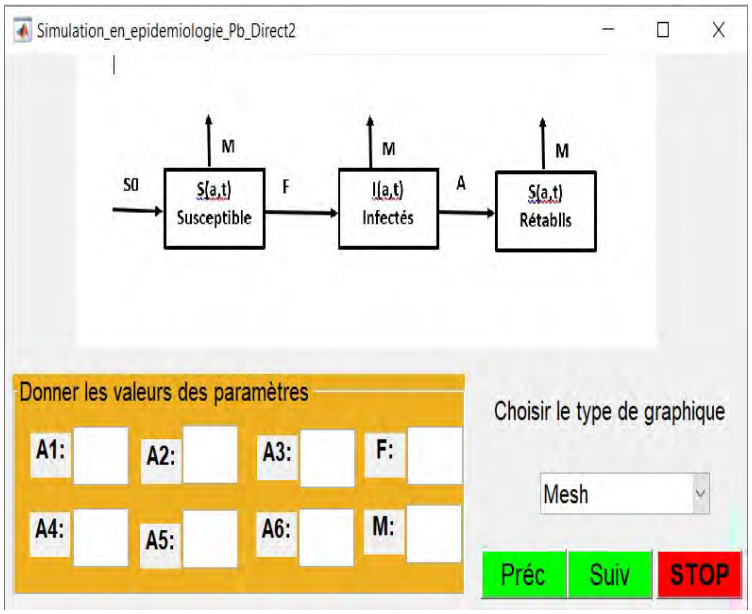


Figure 3.10 – *Interface graphique de résolution de problèmes directs en épidémiologie.*

Ce GUI sert à la résolution de problèmes directs d'un système d'équations issues de la modélisation de la propagation en épidémiologie. Le schéma compartimental présente des paramètres qui peuvent être des coefficients scalaires dans le cas des systèmes linéaires ou des paramètres appartenant à des fonctions de transfert dans le cas des systèmes non linéaires. c'est le cas de (2.1) ou les solutions sont sous forme vectoriel ($u = (s, i, r)$). Dans le cas des systèmes non linéaires, la forme des fonctions vitales est bien connue (Fonction de naissance, mortalité,...), ainsi un programme prend en compte cette structure et il ne reste qu'à intégrer la valeur du paramètre manquant. Le schéma compartimental visible à l'écran, l'utilisateur s'en servira pour donner les valeurs des paramètres manquants, supposés connus, ensuite choisir le type de représentation graphique (*Mesh, Surf, contour, plot*). Les résultats que nous présentons après simulation dans Figure 3.11 sont de la forme vectoriel, $u = (s, i, r)$.

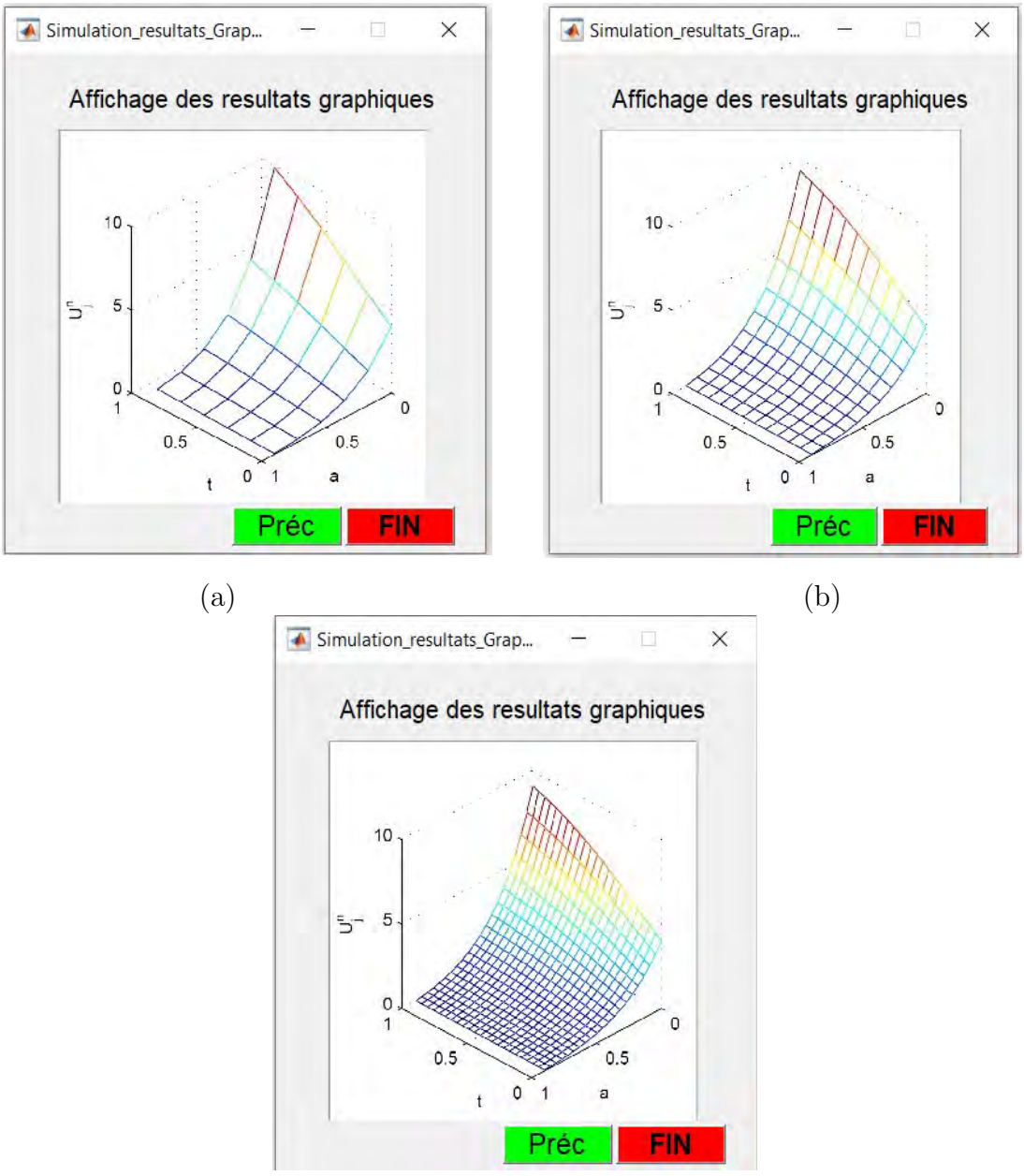


Figure 3.11 – *Résultats de simulation d'un problème direct en épidémiologie obtenus avec le maillage MESH à differents niveaux de discrétisation : $N = 5$(a), $N = 10$(b) et $N = 20$(c).*

Interface(2b(ii)) : Simulation en Epidemiologie Pb Inverse (Figure 3.12)

Il s'agit dans ce GUI de la résolution d'un problème inverse lié au modèle de type SLICRV. Les paramètres contenus dans le modèle compartimental sont supposés inconnus qu'il faut rechercher à partir de données téléchargeables. Ensuite l'opérateur pourra également choisir la méthode de résolution.

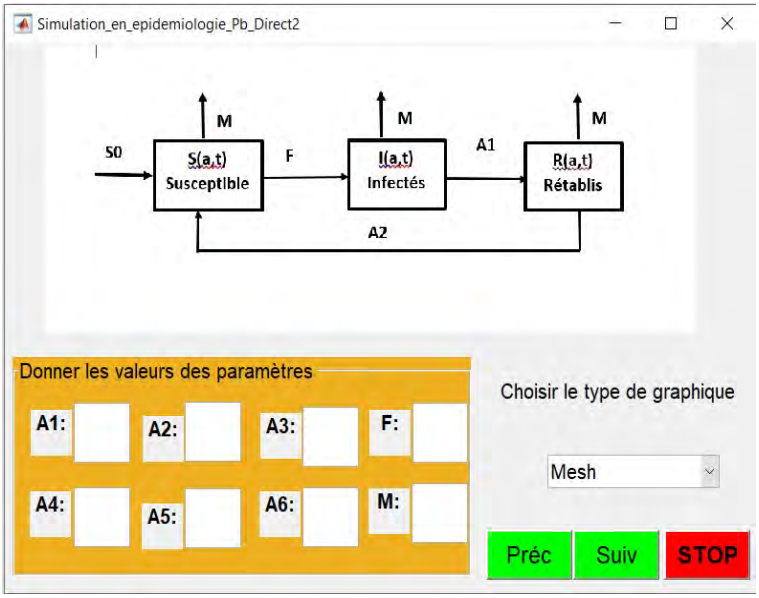


Figure 3.12 – *Interface graphique de résolution de problèmes inverses en épidemiologie.*

Aucun exemple pratique n'a été présenté dans cette section car n'ayant pas été l'objet de publication. Elle pourrait être complétée telque nous l'avions prévu dans la conclusion général.

3.4 Conclusion

Ce GUI que nous présentons dans ce travail est une contribution majeure à la résolution numérique des problèmes issus en environnement et en épidémiologie. Il résume les travaux et les programmes que nous avons développés dans ce mémoire.

Cet outil numérique pourra être enrichi en intégrant de nouvelles interfaces secondaires. Ainsi, en environnement tout comme en épidémiologie, on pourra ajouter de nouvelles interfaces graphiques prenant en compte d'autres modèles mathématiques ou d'autres schémas de résolution numérique.

Conclusion générale

Sommaire

Synthèse des travaux

Ce mémoire fait le bilan d'une dizaine d'années de recherche durant lesquelles nous avons exploré quatre grandes thématiques que sont : La modélisation et la simulation numériques dans les eaux de surface (pollution et assèchement), le maillage des domaines complexes par des éléments finis avec la définition des opérateurs de différentiations, la modélisation et la simulation en épidémiologie et enfin l'analyse asymptotique des équations singulièrement perturbées.

La modélisation et la simulation dans les eaux de surface a longuement été notre axe de recherche. Nous sommes partis de modèles standards d'équations paraboliques avec des conditions aux bords de Dirichlet homogènes pour aboutir à des équations aux dérivées partielles non linéaires avec des conditions aux bords mixtes. La plateforme de simulation s'est beaucoup améliorée avec les nouvelles versions du

logiciel MATLAB. Nos simulations ce sont déroulées sur la version 7.

La nature complexe des domaines d'étude, nous a motivés à proposer des algorithmes de maillage en éléments finis adaptatifs. Cet algorithme génère à la fois les éléments finis triangulaires et les points de collocation dans des proportions calibrées pour avoir une meilleure convergence des solutions en un temps d'exécution relativement faible.

L'épidémiologie est un axe récent dans notre recherche, mais il existe bien une corrélation entre la pollution et des maladies contagieuses. Dans cette thématique nous avons proposé un modèle en se basant sur un existant qui prend en compte une nouvelle forme d'infection. Le calibrage de l'échelle des temps des paramètres en présence nous a conduit à une équation singulièrement perturbée. Nous avons alors appliqué une analyse asymptotique pour approcher la solution. Cette étude a aboutit à un théorème fondamental pour avoir une erreur de l'ordre infiniment petit.

Discussion

Dans l'étude de la modélisation de la propagation de la pollution, nous nous sommes intéressés à l'étude d'une seule espèce de polluant. Pourtant si on considère une décharge de pollution dans l'eau, la nature des polluants est diverse et variée. A cet effet, une modélisation d'un modèle par un système à plusieurs variables prendrait mieux en compte cette diversité. Aussi, les fonctions de diffusion, associées à chaque espèce de polluants auront des échelles de temps différentes (certaine diffusion plus rapide que d'autre). Cela nous conduirait certainement vers des équations de type perturbé, étudiées au chapitre 3, si nous décidions de les ramener avec une même unité de temps. On pourrait dans ce cas envisager une étude asymptotique pour de tel système.

Au niveau des méthodes numériques proposées, nous tirons une satisfaction dans le traitement des domaines complexes par la discrétisation en éléments finis trian-

gulaires et points collocaux. De plus, la série des travaux scientifiques publiés sur la conception des opérateurs de différentiations ont considérablement diminué le temps de calcul machine. Toutefois, avec cette méthode, les tailles des matrices occupent énormément de place dans la mémoire de stockage de la machine. Pour cette raison, un ordinateur ayant une bonne capacité de RAM et une vitesse de processeur assez intéressante contribuera à réduire encore plus le temps de calcul.

Dans la validation des modèles mathématiques développés, surtout dans le cas des problèmes inverses, la collaboration avec des praticiens spécialistes (Ingénieurs, médecins) pourrait aboutir à des modèles performants qui contribueraient à la conception d'une solution numérique commerciale. D'où l'initiative de mettre en place un "graphical user interface" en environnement et en épidémiologie.

Perspectives et évolution

Nous visons après ces travaux de thèse, comme annoncé précédemment, à la mise en place d'un outil numérique qui faciliterait la prise de décision dans l'étude de certains problèmes environnementaux et épidémiologiques. Pour arriver à cet objectif, plusieurs travaux sont en cours notamment :

En environnement :

- L'élaboration d'un modèle mathématique plus raffiné prenant en compte plusieurs espèces polluantes et leurs caractéristiques ;
- Le développement d'un modèle mathématique pour les eaux plus profondes, d'où l'introduction d'une troisième dimension en espace (la profondeur) qui prendrait mieux en compte le flux au niveau de la nappe phréatique.
  Ces deux travaux viendraient généraliser les articles [38, 39, 40] et [46].

En épidémiologie

- Le Développement d'un schéma numérique dans l'approximation de la solution asymptotique. Ensuite faire une comparaison entre les valeurs des erreurs théoriques obtenues dans cette thèse et celles numériques obtenues par le schéma mis en place. Ce travail s'apparenterait au travail de J. BANASIAK et al. [10] et viendrait complèter de manière logique [36].
- Mener des études similaires à [36] dans l'analyse d'autre épidémie telle que le malaria, l'Ebola, etc.

BIBLIOGRAPHIE

[1] M. ABRAMOWITZ, I.A. Stegun (ed.), *Handbook of mathematical functions with formulas, graphs, and mathematical tables,* Wiley-Interscience (1972).

[2] L. M. ABIA and J. C. LOPEZ-MARCOS, *Runge-Kutta methods for age-structured population models*, Appl. Num. Math. 17, pp. 1-17, (*1995*).

[3] L. M. ABIA , O. ANGULO and J. C. LOPEZ-MARCOS, *Age-structured population dynamics models and their numerical solutions*, Ecol. Model. **188** pp. 112–136 (2005).

[4] B.E. AINSEBA, *Exact Controllability, Identifiability, and sentinels.* Ph.D thesis, Compiègne University of technologie, (1992).

[5] B.E. AINSEBA, J.P.KERNEVEZ and R. LUCE, *Application des sentinelles à l'identification des pollutions dans les rivières*, RAIRO, vol. 28, no3, P. 297 à 312, (1994).

[6] R. M. Anderson et R. M. May, *Infectious Diseases of Humans : Dynamics and Control*, Oxford University Press, (1991).

[7] O. ANGULO and J. C. LOPEZ-MARCOS, M. A. LOPEZ-MARCOS and F. A Milner, *A Numerical Method for Nonlinear Age-Structured Population Models with Finite Maximum Age*, J.Math. Anal. Appl. **361** pp. 150–160 (2010).

[8] J. BANASIAK Mathematical Modelling in One Dimension. Cambridge, England : Cambridge University Press, *2013*.

[9] J. BANASIAK, A. TRAORE, S. SHINDIN, R.Y.MASSOUKOU, *Improving Heun's method for solving non linear age-structured population equations with infinite life span*, Pionner journal of advances in applied mathematics, volume 20, Issue 1, pp 55-72 (2017).

[10] J. BANASIAK , A. GOSWAMI, S. SHINDIN. *Aggregation age and space structured population models : an asymptotic analysis approach.* J Evol Eq ; 11 : 121–154 (2011).

[11] J. BANASIAK, M. LACHOWICZ. *Methods of Small Parameter in Mathematical Biology and Others Applications.* Switzerland : Mod Simul Sci Eng Tech, (2014).

[12] J. BANASIAK, W. LAMB. *Coagulation, fragmentation and growth processes in a size structured population.* Disc Cont Dyn Syst ; 17 : 445–472 (2012).

[13] J. BANASIAK, R.Y.M. MASSOUKOU. *A singularly perturbed age structured SIRS model with fast recovery.* Disc Cont Dyn Sys ; 19 : 2383–2399 (2014).

[14] O. BODART, J.P.KERVENEZ and T. MANNIKKO, *Sentinels for Distributed Environment System*, in " Proceedings of the 11th IASTED International Conference on Modelling, Identification and Control",Innsbruck, Austria, February 10-12, (1992).

[15] L. BOS, *On certain configurations of points in Rn which are uniresolvant for polynomial interpolation*, J. Approx. Theory, 64, pp. 271–280 (1991).

[16] L. BOS, M.A. TAYLOR, B.A. WINGATE, *Tensor product Gauss-Lobatto points are Fekete points for the cube*,Math. Comp., 70 , pp. 1543–1547 (2001).

[17] Y. CHA, M. IANNELLI, F. MILLER. *Existence and uniqueness of endemic states for the age-structured SIR epidemic model.* Math Biosci ; 150 : 117–133, (1998).

[18] J.I.DIAZ, J.L. LIONS, *Mathematics, Climate and environment,* Masson Paris Milan Barcelone, (1993).

[19] O. DIEKMANN et J. A. P. HEESTERBEEK, *Mathematical Epidemiology of Infectious Diseases : Model Building, Analysis and Interpretation*, John Wiley & Sons Ltd, (2000).

[20] M. DUBINER, *Spectral methods on triangles and other domains*, J. Sci. Comput.,6, pp.345–390 (1993).

[21] H. FUJIITA, *on the blowing up of solutions of Cauchy problem for* $u_t = \triangle u + u^{1+\alpha}$, J. Fac. Sci. Univ. Tokyo MR, Sect. I 13, pp. 109–124, (1966).

[22] C. GEUZAINE, J.F. REMACLE, *gmsh : a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities*, Int. J. Numer. Meth. Engng, pp. 1–24 (2009).

[23] M. E. Gurtin and R. C. MacCamy, *Nonlinear age-dependent population dynamics*, Arch. Ration. Mech. Anal. **54** 281–300 (1974).

[24] M. GRIEBEL, T. DORNSEIFER, T. NEUNHOEFFER, *Numerical simulation in fluid dynamics. A practical guide*, *SIAM, Philadephia* (1998).

[25] H. W. HETHCOTE, *The mathematics of infectious diseases*, SIAM Review, 42, p. 599–653, (2000).

[26] S. HOWISON *Practical Applied Mathematical Modelling, Analyse, Approximation*, *Cambridge Texts in Applied Mathematics.* (2005).

[27] M. IANNELLI, *Mathematical Theory of Age-Structured Population Dynamics*, Appl. Math. Monographs. C.N.R., Giardini Editori e Stampatori, Pisa (1994).

[28] M. IANNELLI, M. MARTCHEVA and F. A. MILNER, *Gender-Structured Population Modeling : Mathematical Methods, Numerics and Simulations*, SIAM, Philadelphia (2005).

[29] H. A INABA, *semigroup approach to the strong ergodic theorem of the multi state stable population process.* Math Popul Stud ; 1 : 49–77 (1988).

[30] J.P.KERNEVEZ, *the sentinal method and its application to environmental pollution,* CRC Press, Boca raton, FL, (1997).

[31] N. KOSSADOUM, A. TRAORE, N. NGARKODJE, B. MAMPASSI, *On the numerical simulation of Lakes trying-up models*, Far East Journal of applied mathematics Volume79(2), pp 111-126, (2013).

[32] H.A. LEVINE, t*he role of cririca exponents in blowup theorems*, SIAM, 32 , pp 269–288, (1990).

[33] J.L.LIONS, *Sentinelle pour les systèmes distribués à données incomplètes*, Masson, (1992).

[34] J.L.LIONS, *Contrôlabilité exacte,perturbation et stabilisation des systèmes distribués*, tome1, (1998).

[35] J. C. LÒPEZ-MARCOS and J. M. SANZ-SERNA, *Stability and convergence in numerical analysis III : Linear investigation of nonlinear stability*, J. Num. Anal. **8** pp. 71–84 (1988).

[36] R. Y. M. MASSOUKOU1, A. TRAORE, *An age-structured model for the transmission dynamics of hepatitis B : asymptotic analysis*,Turkish Journal of Mathematics, TÜBITAK, 41,pp436-460 (2017).

[37] R. MOSE, M.E.STOECKEL, C.POULARD, P.ACKERER et F.LEHMANN, *Transport parameters identification : application of the sentinel method,* Computational Geosciences 4 pp. 251-273 (2000).

[38] H.NKOUNKOU1, A.TRAORE, M. S. D. HAGGAR and B. MAMPASSI, *Solving Convection Diffusion Problem With a Pseudo Spectral Method on Unstructured Meshes* pioneer Journal of Computer Science and Engineering Technology Volume , N0 1-2, pp. 1-12 (2014).

[39] H. NKOUNKOU, A.TRAORE, G.Seworé, A.Abani and B. Mampassi *Spectral Differentiation on unstructured meshes using jacobi Gauss-Lobatto Points* Far East Journal of Applied Mathematics Volume 59, N0 2, pp. 105-122 (2011).

[40] H. NKOUNKOU, A.TRAORE, G.Seworé, A.Abani and B. Mampassi *Least squares collocation methods for solving partial differential equations : A Matlab appraoch* pioneer Journal of Computer Science and Engineering Technology Volume 1, N0 2, pp. 57-71 (2011).

[41] J.L.LIONS, *Sentinelle pour les systèmes distribués à données incomplètes*, Masson, (1992).

[42] R. PICHS-MADRUGA, Y. SOKONA, E. FARAHANI, S. KADNER, K. SEYBOTH, A. ADLER, I. BAUM, S. BRUNNER, P. EICKEMEIER, B., KRIEMANN JS, S. SCHLÖMER, C. VON STECHOW, T. ZWICKEL and J.C. MINX editors. *Climate Change 2014, Mitigation of Climate Change Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA. : Cambridge University Press ; *2014*.

[43] K. TANYA, *An Explicit Third-order numerical Method for Size-Structured Population Equations*, Num. Meth. PDE's **19 (1)** pp. 1-21 (2002).

[44] M. A. TAYLOR, B. A. WINGATE, and R. E. VINCENT, *An algorithm for computing FEKETE points in the triangle*, SIAM,38, pp. 1707–1720 (2000).

[45] Thieme HR. Mathematics in Population Biology. Princeton, NJ, USA : Princeton University Press, 2011.

[46] A. TRAORE, B. MAMPASSI, *Least squares spectral collocation method for solving identification problems in a Lake pollution model over a complex domain* , International Journal of Mathematical and Computational Methods Volume 2, (2017).

[47] A. TRAORE, B. MAMPASSI, L. SOME, *A least-squares spectral collocation formulation for solving PDEs on complex geometry domains*, International Journal of Applied Mathematics and Computation Volume 2 no. 4, pp 9-22,(2011).

[48] A. TRAORE, *Contribution à la résolution numérique de problèmes de détection de pollution en milieu fluide à structure géométrique complexe*, thèse de troisième cycle (2008).

[49] A. TRAORE, B. MAMPASSI, B. SALEY, *A Numerical Approach of the sentinel method for distributed parameter systems, CEJM(5)4 ,* pp. 751-763 (2007).

[50] UNESCO, *L'eau et la santé à l'occasion de la journée mondiale de la santé*, bulletin d'information du portail de l'eau de l'UNESCO, no. 87 pp. 1-8, (2005).

[51] T.C. WARBURTON, S.J.SHERWIN, and G.E. KARNIADAKIS, *Basis Functions for Triangular and Quadrilateral High-Order Elements*, SIAM, 20, pp. 1671–1695 (1999).

[52] G. F. WEBB, Theory of Age Dependent Population Dynamics, Marcel Dekker, New York (1985).

[53] D. ZIRIRANE, J.J.BAGALWA, M. ISUMBISHO, M. MULENGEZI, I. MUKUMBA et al. *Evolution comparée de la pollution des rivières Kahuwa et Ppungwe par l'utilisation des macroinvertébrés benthiques*, VertigO-la revue électronique en sciences de l'environnement, volume 14 no3, (2014).

[54] L. ZOU, S. RUAN, W. ZHANG. *Modeling the transmission dynamics and control of hepatitis B virus in China.* J Theor Biol ; 262 : pp. 330–338 ; (2010).

[55] L. ZOU, S. RUAN, W. ZHANG *An age-structured model for transmission dynamics of hepatitis B.* SIAM J Appl Math ; 70 : pp. 3121-3139 (*2010*).

ANNEXES

ANNEXE 1

Article :

A. TRAORE, B. MAMPASSI, *Least squares spectral collocation method for solving identification problems in a Lake pollution model over a complex domain*, International Journal of Mathematical and Computational Methods Volume 2, pp 19—30, (2017)

Least squares spectral collocation method for solving identification problems in a Lake pollution model over a complex domain

ABOUBAKARI TRAORE
Ecole normale superieure
Department of mathematics
08 BP 10 ABIDJAN
COTE D'IVOIRE
traboubakari@yahoo.fr

BENJAMIN MAMPASSI
Cheikh Anta Diop University
Department of mathematics
BP 5005 Fann, DAKAR
SENEGAL
mampassi@yahoo.fr

Abstract: In this paper, We modeled the behavior of pollution concentration in a lake by a parabolic equation. The domain of the lake is reduced to 2D-dimension in space and characterized by some obstacles inside and the circumference is a polygonal form. First, The mathematical model obtained contains unknown parameters which have to be determined and then the approximative solution of the mathematical model has to be estmimated. For this purpose, we approximate the system of equations by means of discrete differential operators adapted to the complexity of the domain based on Least Squares Spectral Collocation Method, LSSCM. To test our numerical scheme, we consider some experimental data. After computations, we obtain optimals values of unknown parameters and the approximative solution of the mathematical model. The 2D and 3D-dimension in space graphics of the solution are presented. The analysis of the graphs allows us to better identify the source of the pollution, the concentration of the pollution and the direction of the propagation in lake. We conclude that the use of Least squares spectral collocation method to solve pollution problem over a complex domain was successful.

Key–Words: Lake Pollution, Least-squares formulation, finite elements, Collocation point methods, differentiation matrices, complex domain

1 Introduction

In this paper, we describe the propagation of pollution concentration in a lake by a non-linear partial differentiation equation. This mathematical model is acheived by taking into account the physical, chemical and biological properties of water. The system of equations obtained contain some unknown parameters coming from the modelling process.

The mathematical problem we are facing is: can we recover both unknown parameters and the solution of the system equation (if there exists) knowing some measurements of pollution concentration in a subset of the lake? this is an identification problem. A such problem has been investigated by many autors both theoretical and numerical aspects and has been applyed in diverse fields of reseach; see [19] and[24] for more details. For an application in physic, see [7] and [22], the authors work on identification problem arising in single-photon emission computerized tomography. In identification problems, the measurements and where they have been collected play an important role so in [3], [1] and [2], the determination of some characteristic sources is based on boundary data. But there are restrictions on the number and type of sources that can be identified from boundary data. For more information see [2]. In some articles, [4], [5], [6] and [8], the authors present numerical methods to identify the source terms in a model of pollution propagation in surface water. This topic is close to our present research. But we do not limit the identification problem to the source term.

Two most popular methods are often used in identification problems: the least squares method (see [9], [16] and [30]) and the sentinel method (see [21], [23] and [27]). In many papers, both numerical and theoretical aspects for these methods are developed using smooth domains.

This paper deals with a computational method for determining unknown terms of a non linear partial differential equation including unknown parameters. Theses unknown terms may be in initial, boundary conditions or source terms. Furthermore, to be close to the reality, we assume that the computational domain has a complex geometry shape (see Figure 1). We intend to take into account the geometry complexity as well as the non linear term from which blow up property can occur. In this case, the computation of PDEs requires a special treatment both for meshing domain and the discretization of the equation system. To this end, we use LSSCM over a complex domain

in the current work.The most important papers where LSSCM has been developed are given by [13], [15], [17] and [26]. This paper extends a previous work, [26], by solving both a parabolic equation and identification problem.
This paper is organized as follow: In section 2, we present a mathematical model of the pollution propagation and the description of the domain over which the problem is posed. Then the least squares spectral collocation formulation of parameters identification problem is posed. In section 3, we present the discretization process of both system equations and the domain. In the section 4, we present some numerical results and give some remarks in the last section.

2 Setting of the problem

The polluted lakes generally contain chemical wastes like nitrate and phosphate coming from industries, agricultural runoff and waste water from cities. These pollutants kill fishes and other aquatic animals by reducing the rate of dissolved oxygen in the water. So, there exists a correlation between the rate of dissolved oxygen and the rate of pollution in the water and that is why the quantity of pollution is valued from the quantity of dissolved oxygen that pollutants need for their chemical and biological reactions. This quantity is estimated in BOD (Biologic Oxygen Demand) and COD (Chemical Oxygen Demand) respectively (see [12] and [18]). We suppose in the remaining of this paper that $C(t,\mathbf{x})$ (Kg/m^3) denotes the concentration of the pollution in the Lake measured in COD at the time t and $\mathbf{x}$-position.

To take into account the fluid properties, we describe the pollutant propagation by considering its concentration $C(t,\mathbf{x})$. According to some assumptions the following terms will be considered to describe our model:

- A diffusion term, $k.div\left[a(\mathbf{x})\nabla C(t,\mathbf{x})\right]$, where k is a constant, $a(\mathbf{x})$ denotes the diffusion of chemical substances in the water. In the case of the Lake, $a(\mathbf{x})$ may be considered as a constant term so that the diffusion term becomes $K.\Delta C(t,\mathbf{x})$. K is the coefficient of diffusion (m^2/s).
- A transport term, $\overrightarrow{u}\nabla C(t,\mathbf{x}),$ where $\overrightarrow{u}$ denoted the velocity of the fluid. In absence of wind, considered as the only means of transportation here, we can assumed that $\overrightarrow{u}=\overrightarrow{0}$.
- A reaction term, $\lambda C(t,\mathbf{x})-\mu\left|C\right|^p(t,\mathbf{x})$, where λ and μ are coefficients that describe the characteristics of the reaction process. It shows chemical and biochemical interactions in the fluid so two types of reactions may be considered: the first term increases the rate of pollution and the second term reduces the pollution rate.
- A source term $f(t,\mathbf{x})$. It brings polluted substances in the liquid. We have two cases of source term. The pollutants which come from land surface (agricultural runoff, waste water) denoted by $\xi(t,\mathbf{x})$ and those which are situated at the bottom of the Lake (sediment, Heavy metals). Let consider we have N_s sources terms at $\mathbf{x}_i-$ coordinate. So, the general formulation of the source can be taken like

$$f(t,\mathbf{x})=\xi(t,\mathbf{x})+\sum_{i}^{N_s}\lambda_i\widehat{\xi_i}(t)\times\delta(\mathbf{x}-\mathbf{x}_i),$$

where $\delta(\mathbf{x}-\mathbf{x}_i)$ represents the Dirac function associated to $\mathbf{x}_i$.

The domain considered in this problem has two types of boundary: Γ_{int}, the circomference of obstacles inside the domain and Γ_{out}, the limit of Lake domain. Then the boundary of the domain, $\Gamma=\Gamma_{int}\cup\Gamma_{out}$ (see Figure 1). For the sake of simplicity, we suppose that there are no exchange of pollution concentration through the boundaries. So, we are concerned with a Dirichlet condition

$$C\left|_{\Gamma}=0\right.$$

and the initial condition is

$$C(0,\mathbf{x})=g(\mathbf{x};\tau).$$

We summary the behavior of pollution concentration C in a lake by the parabolic system below

$$\begin{aligned}
\frac{\partial C(t,\mathbf{x})}{\partial t} &= \xi\Delta C(t,\mathbf{x})+\eta C(t,\mathbf{x}))-\mu\left|C\right|^p(t,\mathbf{x})) \\
&\quad +f(t,\mathbf{x};\lambda), \qquad (t,\mathbf{x})\in]0,T]\times\Omega, \\
C(t,\mathbf{x}) &= 0, \qquad (t,\mathbf{x})\in]0,T]\times\Gamma, \\
C(0,\mathbf{x}) &= g(\mathbf{x};\tau), \qquad \mathbf{x}\in\Omega
\end{aligned} \tag{1}$$

where

- $\Omega\subset\mathbb{R}^2$ and Γ is the boundary of Ω.
- $]0,T]$, $T>0$, is the spending time for the experience.
- The reals ξ,η,μ,p,λ and τ are unknown parameters and strictly positives.

- The structure of $f(t, \mathbf{x}; \lambda)$, the source of pollution and $g(t, \mathbf{x}; \tau)$, the initial concentration are known.

We are concerned with the identification problem that consists of determining the unknown parameters ξ, η, μ, p, λ and τ in the system (1). For theoretical aspects, on can refer to [14], [20] and [21] for the existence and uniqueness conditions of global solution for the system (1).

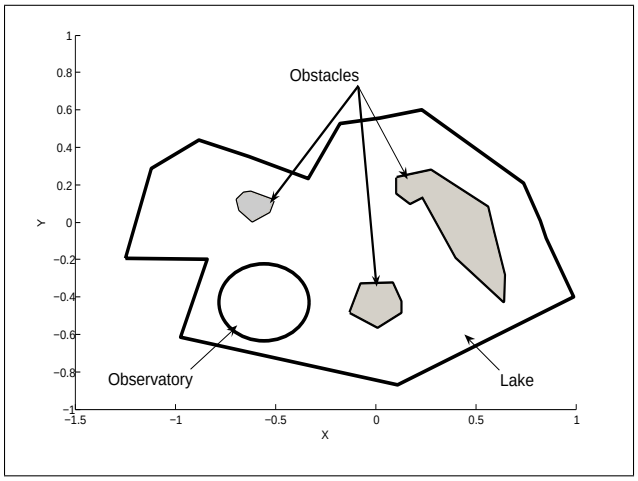


Figure 1: *The domain of the Lake.*

3 Discrete differential operators

The solving of (1) requires accurate approximation of derivatives and particularly on such domain. Least squares spectral collocation method combines both the standard least squares method and collocation method. For this purpose, a macro mesh is used to mesh the whole domain. LSSCM consists in solving (1) into each triangular finite elements and the global solution is obtain by applying the standard least squares method after the assembly process. The convergence of LSSCM has been hugely study by previous authors, we mentioned at the beginning, in smooth domain. The case of complex domains has been studied in [26] for elliptic equations. To better understand this method, we describe first the discrete differentiation matrix associated with Fekete([25]) or Gauss-Lobatto([29]) points in the triangles.

3.1 Discrete differential operators

Let us consider as standard triangle the following

$$\widehat{\Theta} = \{(r, s), -1 \leq r, s \leq 1;\ r + s \leq 0\}. \quad (2)$$

We denote by $p_j^{\alpha,\beta}(s)$ the Jacobi polynomials of (α, β)-order and degree j [**?**]. It is well known these polynomial have $j + 1$ zeros and are distributed arbitrary over $]-1, 1[$. In the case of standard quadrangle, the Gauss-Lobatto points are generated using tensor product (see [28]). In the case of a triangle, one can build a suitable transformation function to translate the Gauss-Lobatto points from a quadrangle to a standard triangle [29]. It is proved in [10] and [11] that these collocation points are close to the Fekete ones. Let consider $P_N(\widehat{\Theta})$ the space of polynomials of degree less than N over $\widehat{\Theta}$. In the remaining of this paper we shall by note $\{\phi_k\}_{1 \leq k \leq N}$ a basis of $P_N(\widehat{\Theta})$.

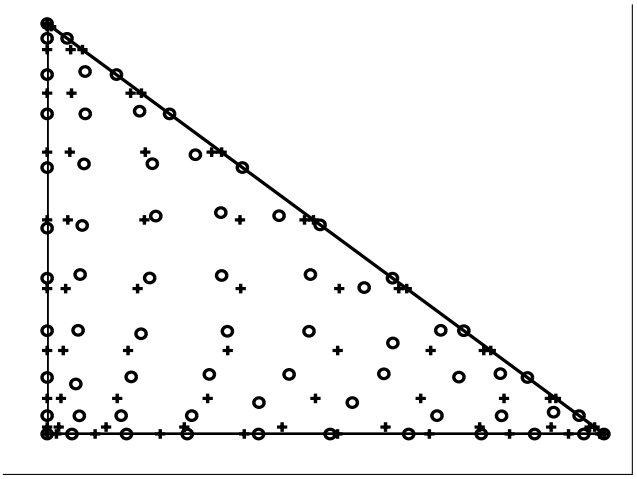


Figure 2: *The Lobatto triangle nodes (+) and associated Fekete nodes (o) over the reference triangle $\widehat{\Theta}$.*

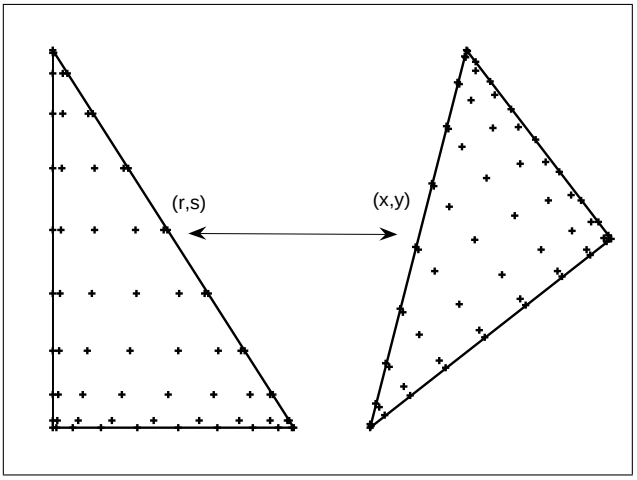


Figure 3: *Distribution of collocation points from reference triangle (left) to arbitrary triangle (right).*

3.2 Differentiation over the reference triangle

For any continuous function $u(r, s)$ on the reference triangle $\widehat{\Theta}$, we can write its spectral approximation in

the space $P_N(\Theta)$ by

$$u^N(r,s) = \sum_{k=1}^{N} U_k \phi_k(r,s) \tag{3}$$

where the coefficients U_k are obtained using collocation equations at collocation points that we denote by $\widehat{z}_m(r,s)$:

$$u^N(\widehat{z}_m) = \sum_{k=1}^{N} U_k \phi_k(\widehat{z}_m), \quad m = 1, 2, ..., N. \tag{4}$$

Setting $U(t) = \left(u^N(\widehat{z}_1),\ u^N(\widehat{z}_2),\ ...,\ u^N(\widehat{z}_N)\right)'$ and $C(t) = (U_1(t),\ U_2(t),\ ...,\ U_N(t))'$ the coefficient vectors, where $(..)'$ designs the transposed vector. The equation yields

$$C = V^{-1} \times U \tag{5}$$

where V is the Vandermonde matrix over collocation points. That is a matrix whose components are $\phi_k(\widehat{z}_m)$. According to (3), the derivatives in s and in r directions at collocation points $\widehat{z}_m$ are then given by

$$\partial_r u^N(\widehat{z}_m) = \sum_{k=1}^{N} U_k \times \partial_r \phi_k(\widehat{z}_m)$$

and

$$\partial_s u^N(\widehat{z}_m) = \sum_{k=1}^{N} U_k \times \partial_s \phi_k(\widehat{z}_m)$$

respectively. We introduce two differentiation matrices V^r and V^s, of the size $N \times N$ respectively in $r-$ and $s-$direction respectively, whose components are $V^r_{ij} = \partial_r \phi_j(\widehat{z}_i)$ and $V^s_{ij} = \partial_s \phi_j(\widehat{z}_i)$ respectively. Denoting U_r and U_s, the vector values of the differential approximations in $r-$ and $s-$ direction respectively at collocation points, we obtain

$$U_r = D^r \times U \quad \text{and} \quad U_s = D^s \times U \tag{6}$$

where we have set

$$D^r = V^r \times V^{-1} \text{ and } D^s = V^s \times V^{-1}. \tag{7}$$

3.3 Differentiation over an arbitrary triangle

Any derivative over an arbitrary triangle is derived from the reference triangle according to the bijective transformation such that:

$$u(r,s) = u(x(r,s), y(r,s)). \tag{8}$$

Applying the derivative rule in r (respectively s) direction, we have

$$\begin{cases} \partial_r u &= (\partial_r x)\,\partial_x u + (\partial_r y)\,\partial_y u, \\ \partial_s u &= (\partial_s x)\,\partial_x u + (\partial_s y)\,\partial_y u. \end{cases} \tag{9}$$

Let us denote by $U_x(t)$ and $U_y(t)$ the vector values of the differential approximation in x and y directions respectively at collocation points. Then, from (7) and (9) we deduce

$$\begin{pmatrix} U_x \\ U_y \end{pmatrix} = G^{-1} \times \begin{pmatrix} D^r \\ D^s \end{pmatrix} \times U \tag{10}$$

where the matrix G is associated to the system (9) over collocation points. Setting $\mathbb{D} = G^{-1} \times \begin{pmatrix} D^r \\ D^s \end{pmatrix}$ then the differentiation matrices over an arbitrary triangle in $x-$direction, and $y-$direction are obtained by extracting the matrix $\mathbb{D}$ from the first to the N^{th} column and from the $(N+1)^{th}$ to $2N^{th}$column respectively :

$$D^x = \mathbb{D}(1:N,\ :\ ) \quad \text{and} \quad D^y = \mathbb{D}(N{+}1:2N,\ :\ ). \tag{11}$$

The second order differentiation matrices on an arbitrary triangle are obtained in the similar way. Next we shall denote by D^{xx} and D^{yy} the second differentiation matrix in $x-$direction and $y-$direction respectively and by $\mathbb{L}$ the discrete Laplacian, where

$$\mathbb{L} = D^{xx} + D^{yy}. \tag{12}$$

3.4 Differentiation over the entire domain

We are concerned in this paper to $2D-$ space and the triangular finite elements mesh. To define the differentiation matrix over the whole domain like the previous sections, we explain briefly the collocation points assembly process. Let denote by $\{\Theta_k\}_{k=1,...,N_\Theta}$ the set of elementary triangle in the domain of length N_Θ and $\mathbf{X}_{loc}$ the set of collocation points coming from each Θ_k of length N_{loc}. It's clear that $\mathbf{X}_{loc}$ contains some collocation points repeated more than ones, like certain points on edges and nodes (see Figure 4). To avoid the repeated points, we achieve a second set, $\mathbf{X}_{glob}$, of length N_{glob} ($N_{loc} > N_{glob}$). We have the following relation

$$\mathbf{X}_{glob} = \mathbb{Z} \times \mathbf{X}_{loc}$$

where $\mathbb{Z}$ is the assembly matrix of size $N_{glob} \times N_{loc}$. A such matrix were defined in [29] for hybrid elements(triangular and quadrilateral elements). The components of $\mathbb{Z}$ are 0 or 1.

Let U_{glob} the vector solution at $\mathbf{X}_{glob}$ collocation points.

$$U_{glob} = \left(u_1, u_2, ..., u_{N_{glob}}\right)'.$$

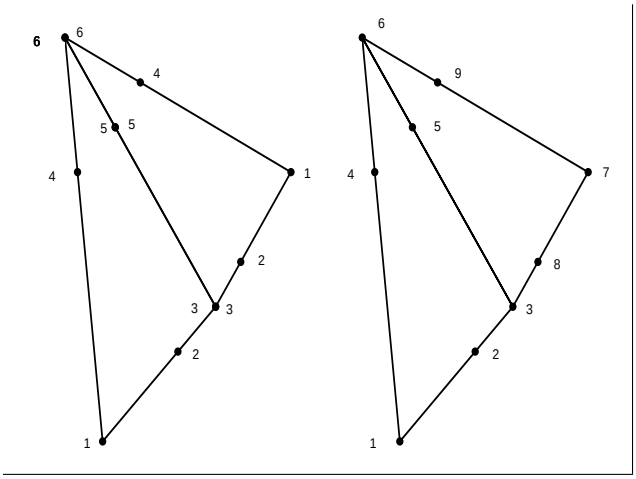


Figure 4: *(left) Local numbering of two elementary triangles, the size of $\boldsymbol{X}_{loc}$ is twelve(12) collocation points per triangle. (right) Global numbering of two elementary triangles, the size of $\boldsymbol{X}_{glob}$ is nine(9) collocation points.*

We define the first derivative matrix in x-direction of U_{glob} over the whole domain by the following relation:

$$\partial_x U = \mathbb{H}_x \times \mathbb{Z} \times U_{glob}$$

where $\mathbb{H}_x$ is a diagonal matrix of size $d(N)^2 = (N \times N_\Theta)^2$

$$\mathbb{H}_x = \begin{pmatrix} D_x^1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & D_x^{N_\Theta} \end{pmatrix} \tag{13}$$

where D_x^k represent the local first derivative matrix in x-direction over an elementary triangle Θ_k. Then we deduce the global differentiation matrix in x-direction, given by:

$$\mathbb{D}_x = \mathbb{H}_x \times \mathbb{Z}. \tag{14}$$

In the same ways, we define below the first derivative matrix in y-direction and the second derivative matrix in x- and y-direction respectively over the whole domain by:

$$\mathbb{D}_y = \mathbb{H}_y \times \mathbb{Z}; \tag{15}$$

$$\mathbb{D}_{xx} = \mathbb{H}_{xx} \times \mathbb{Z}; \tag{16}$$

and

$$\mathbb{D}_{yy} = \mathbb{H}_{yy} \times \mathbb{Z} \tag{17}$$

where $\mathbb{H}_y$, $\mathbb{H}_{xx}$ and $\mathbb{H}_{yy}$ are like (13) where D_x^i are replaced by D_y^i, D_{xx}^i and D_{yy}^i respectively , $i = 1, ..., N_\Theta$.

4 Least squares spectral collocation formulation

Using the discretization method explain above, the discrete system of (1) can be written as follow

$$\boxed{\begin{array}{rcl} \dfrac{d\underline{C}(t)}{dt} - \xi \mathbb{L} \times \underline{C}(t) + \eta \underline{C}(t) - \mu \left|\underline{C}\right|^p (t) & = & \underline{f}(t;\lambda) \\ D_b \times \underline{C}(t) & = & 0 \\ \underline{C}(0) & = & \underline{g}(t;\tau) \end{array}} \tag{18}$$

where $t \in]0, T]$,

$$\underline{C}(t) = \left(C(t, \mathbf{x}_1), ..., C(t, \mathbf{x}_{d(N)})\right)', \tag{19}$$

$$\underline{f}(t;\lambda) = \left(f(t, \mathbf{x}_1; \lambda), ..., f(t, \mathbf{x}_{d(N)}; \lambda)\right)', \tag{20}$$

$$\underline{g}(t;\tau) = \left(g(t, \mathbf{x}_1; \tau), ..., g(t, \mathbf{x}_{d(N)}; \tau)\right)', \tag{21}$$

the $\mathbf{x}_k$, $k = 1, ..., d(N)$ are the collocation points. D_b designs the discrete operator which selects the boundary points. The residual of the discrete system (18) is valued by the following operator

$$\begin{array}{rcl} \mathscr{L}(\underline{v}, \underline{C}) & = & \left\| \dfrac{d\underline{C}(t)}{dt} - \xi \mathbb{L} \times \underline{C}(t) + \eta \underline{C}(t) \right. \\ & & \left. -\mu \left|\underline{C}\right|^p (t) - \underline{f}(t;\lambda) \right\|^2. \end{array} \tag{22}$$

Let denote the vector of unknown parameters by

$$\underline{v} = (\xi, \eta, \mu, p, \lambda, \tau). \tag{23}$$

To estimate $\underline{v}$ we need some data measurements of C which govern the system (1) over the whole domain Ω. It seems to be difficult in reality to get such data. In this paper, we are concerned with the following type of inverse problem: Having some data about the solution over a sudomain of Ω, we want to estimate $\underline{v}$ and the numerical solution $\underline{C}(t)$ in the entire domain. To solve this problem we have been inspired by the sentinel method developped by J.L. Lions [21]. We denote by C_{obs} the so called data and $\Omega_{obs} \subset \Omega$ the observatory, a subdomain of Ω (Figure 6) where the measurement has been done. We denote by D_{obs} and D_b the matrix which select the observatory respectively the boundary collocation points from whole collocation points of Ω such that

$$\begin{array}{rcl} D_{obs} \times \underline{C} & = & C|_{x \in \Omega_{obs}} \\ D_b \times \underline{C} & = & C|_{x \in \Gamma}. \end{array} \tag{24}$$

We are concerned here with the determination of $\underline{C}(t, \underline{v})$ solution of (18) and the unknown vector (23).

We summary the least squares spectral collocation method of (18) as

$$\boxed{\begin{array}{c}\text{find } \underline{v}^*, \underline{C}^* \text{ solutions of} \\ J_{\widehat{\beta}}(\underline{v}^*, \underline{C}^*) = \min\limits_{(\underline{v},\underline{C})\in\mathbb{R}^6\times\mathbb{R}^{d(N)}} J_{\widehat{\beta}}(\underline{v}, \underline{C})\end{array}} \quad (25)$$

where

$$\begin{aligned} J_{\widehat{\beta}}(\underline{v}, \underline{C}) &= \sum_{k=1}^{[T/\Delta t]} \|D_{obs} \times \underline{C}(k\Delta t) - \underline{C}_{obs}(k\Delta t)\|^2 \\ &+ \sum_{k=1}^{[T/\Delta t]} \|D_b \times \underline{C}(k\Delta t)\|^2 + \widehat{\beta} \times \|\underline{v} - \widehat{\underline{v}}\|^2 \\ &+ \mathscr{L}(\underline{v}, \underline{C}) \end{aligned} \quad (26)$$

and $\underline{C}_{obs}(t)$ is a measurement vector obtain at $\mathbf{x} \in \Omega_{obs}$ at time t. A positive constraint is established on $\underline{C}$. $\widehat{\beta}$ stand for Tikhonov regularization parameter and $\widehat{\underline{v}}$ is an a priori information, (27). We look for values of $\underline{v} \in]0, 2]$ according to this information

$$\boxed{\begin{array}{ccccc} 0 & < & \xi & < & 10^{-3} \\ 10^{-6} & < & \eta & < & 0,4 \\ 0,1 & < & \mu & < & 6 \\ 0,034 & < & p & < & 2 \\ 10^{-6} & < & \lambda & < & 0,5 \\ 10^{-2} & < & \tau & < & 1,6 \end{array}} \quad (27)$$

We summary the computational process through an algorithmic scheme (Figure 5).

5 Numerical experimentation

Collecting Data: For experimental purpose to test our numerical scheme, we built data, $\underline{C}_{obs}(t)$ as:

Figure 8: firstly, We consider graphs of pollution at five arbitrary points, O_i, $i = 1, ..., 5$ (Observation points, Figure 6) situated in Ω_{obs},

Figure 7: secondly, we collect the value of each curve at 24 axis points representing the time intervals (per days) and

finally we fit the data over the whole collocation points include to Ω_{Obs}.
The expression of source function f is given by

$$f(t, \mathbf{x}) = 10\lambda sin(\pi t + 1)\delta_{\mathbf{x}_c}(\mathbf{x})$$

where $\delta_{\mathbf{x}_c}(\mathbf{x})$ is a Dirac function define in $\mathbf{x}_c$, here $\mathbf{x}_c = (0, 0)$, such that

$$\delta_{\mathbf{x}_c}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{x}_c \\ 0 & \text{otherwise.} \end{cases}$$

The initial function is given by

$$g(x) = \tau.\left(x^2 + y^2 - .06\right)\delta_{\mathbf{x}_c}(\mathbf{x}).$$

We use MATLAB 7.01 graphic interface to compute the algorithm which has some solvers associated to our numerical scheme. So, for the achievement of (25) we used " *fmincon*" solver and for (22) we used the "*ode45*" solver.

Results: By using the data built in Figure 7, for $0 \leq t \leq 24$ and after the computation of (25), the optimal values of $\underline{v}$ obtained are:

$$\boxed{\begin{array}{rc} \xi = & 58.10^{-4} \\ \eta = & 10^{-6} \\ \mu = & 6 \\ p = & 0.3373 \\ \lambda = & 0.4999 \\ \tau = & 1.4680 \end{array}}$$

The approximative values of the parameters are the main results of the numerical process. Its identify clearly the parabolic equation (1):
The expression of source function f is determined

$$f(t, \mathbf{x}) = 10\lambda sin(\pi t + 1)\delta_{\mathbf{x}_c}(\mathbf{x}) \quad \textit{with} \quad \lambda = 0.4999 \quad (28)$$

the initial function is identified

$$g(x) = \tau.\left(x^2 + y^2 - .06\right)\delta_{\mathbf{x}_c}(\mathbf{x}) \quad \textit{with} \quad \tau = 1.4680$$

and the system of equation which governed our pollution model is dertermined:

$$\begin{aligned} \frac{\partial C(t, \mathbf{x})}{\partial t} &= \xi\Delta C(t, \mathbf{x}) + \eta C(t, \mathbf{x})) - \mu\left|C\right|^p(t, \mathbf{x})) \\ &+ f(t, \mathbf{x}; \lambda), \qquad (t, \mathbf{x}) \in]0, T] \times \Omega, \\ C(t, \mathbf{x}) &= 0, \qquad (t, \mathbf{x}) \in]0, T] \times \Gamma, \end{aligned} \quad (29)$$

with

$$\xi = 58.10^{-4}; \quad \eta = 10^{-6}; \quad \mu = 6 \quad \textit{and} \quad p = 0.3373. \quad (30)$$

Then one can approximate the solution of (29) by any suitable partial differential equation solver. The approximative solution is given by $\underline{C}^*$.
For time goes from 0 to 20, we plot in two dimensions (Figures 10 and 11) and in three dimensions (Figure 9) the graphs of the approximative solution $\underline{C}^*$ so that we can appreciate the evolution of the experimental solution. We have selected four pictures at arbitrary time ($t = 2$, $t = 4$, $t = 6$ and $t = 20$). So in:

Figure 9: Firsly, We observe the growing of the volume of the concentration when the time of the experience is increasing. Secondly, the shape of the solution contains some holes located at the same positions of our obstacles introduced in the domain. That means, there are no pollutions in these places and it confirms our hypotheses. The groth of $\underline{C}^*$ is consistent the graph of data (Figure 8). Finally, at each time, we remark somewhere in the graph a highest value of the pollution concentration, it shows the position of the source term as predicted in 28.

Figure 10 and 11: we decide to plot in $2D-$ dimension to point out the value of $\underline{C}^*$ around the boundaries. When the time increases, We remark that the value of pollution concentration around the boundaries(circumference, obstacles) increase. In particular, in Figure 10 there is an accumulation of streamline around the boundaries. In Figure 11, the color of the domain moves from blue ($t = 2$) to green ($t = 20$). the source term position is identified by the heighest value, the red color and The surfaces of obstacles are maintained in blue color, the lowest value. Here again, the source term is clearly identify and the growth of pollution is in phase with the previous graphs.

Figure 12: we are interested by the direction of the propagation of the pollution, so we have presented the gradient of $\underline{C}^*$ in $2D-$ dimension. The arrows show the senses of the propagation in the domain. When the time increases the arrows move from the source position throughout the entire domain excepted in the obstacles. Also, most of the arrows are directed to the boundaries. That confirms the accumulation of pollution around the boundaries in Figure 10.

6 Concluding remarks

In this paper, we acheive a convenient algorithm to study the propagation of pollutants in a lake over a complex domain. The succes of the numerical scheme is done by coupling the technique of differentiation operators over finite elements and the least squares spectral collocation method. One particularity of this algoritm, it does not need the data over the entire domain. It works with few data collected in a subdomain of the lake like in sentinel method [21]. The numerical test with experimental data gives an excellent results that identify clearly the mathematical model, localyze the source terme and the direction of pollutants. the data are consistence with our results obtained.

References:

[1] C. J. ALVES, N. F. MARTINS and N. C. ROBERTY, Full identification of acoustic sources with multiple frequencies and boundary measurements, *Inverse Problems and Imaging*,3, 2009, pp. 275–294.

[2] C. J. ALVES, R. MAMUD, N.F.M. MARTIN and N. C. ROBERTY, On inverse problems for characteristic sources in Helmholtz equations, *Hindawi*,ID 2472060,2017, 16 pages.

[3] S. ACOSTA, S. CHOW, J. TAYLOR, and V. VILLAMIZAR, On the multi-frequency inverse source problem in heterogeneous media, *Inverse Problems*, 28, 2012.

[4] M. ANDRLE and A. EL BADIA, On an inverse source problem for the heat equation. Application to a pollution detection problem, II, *Inverse Problems in Science and Engineering*,23, 2015, pp.389–412.

[5] A. EL BADIA and T. HA-DUONG, On an inverse source problem for the heat equation. Application to a pollution detection problem, *Journal of Inverse and Ill-Posed Problems*, 10, 2003, pp.585–599.

[6] A. EL BADIA, T. Ha-DUONG and A. HAMDI, Identification of source in a linear advection-dispersion-reaction equation: application to a pollution source problem, *Inverse problems*,21, 2005, pp.1–7.

[7] G. BAL, A. JOLLIVET, Combined source and attenuation reconstructions in SPECT, in Tomography and Inverse transport Theory, *Contemp. Math., AMS, Providence, RI,* 559, 2011, pp. 13–27.

[8] A. E. BAPTISTA, E. E. ADAMS, and K. D. STOLZENBACH, Eulerian-Lagrangian Analysis of Pollutant Transport in Shallow Water, *Ralph M. Parsons Laboratory, MIT, Rpt.*, 296, 1984, (2004).

[9] P. BOCHEV and M. GUNZBURGER, least squares finite elmement methods, *Prodeeding of the International Congress of Mathematicians*,3,2006, pp. 1137–1162.

[10] L. BOS, On certain configurations of points in Rn which are uniresolvant for polynomial interpolation, *J. Approx. Theory*, 64, 1991, pp. 271–280.

[11] L. BOS, M.A. TAYLOR, B.A. WINGATE, Tensor product Gauss-Lobatto points are Fekete points for the cube, *Math. Comp.*, 70 , 2001, pp. 1543–1547.

[12] J.I.DIAZ, J.L. LIONS, *Mathematics, Climate and environment,* Masson Paris Milan Barcelone, 1993.

[13] M. FERNANDINO, C. A. DORAO, The least squares spectral element method for the Cahn-Hilliard equation *Elselvier, Applied mathematical modelling*, 35, 2011, pp. 797–806.

[14] H. FUJIITA, on the blowing up of solutions of Cauchy problem for $u_t = \triangle u + u^{1+\alpha}$, *J. Fac. Sci. Univ. Tokyo MR*, Sect. I 13, 1966, pp. 109–124.

[15] W. HEINRICHS, Least-Squares Spectral Collocation with the Overlapping Schwarz Method for the Incompressible Navier–Stokes Equations, *Springer, Numerical Algorithms*,43, 2006, pp. 61–73.

[16] T. KARIYA, H. KURUTA, *John and Son, Ltd*, 2004.

[17] T.KATTELANS, W. HEINRICHS, A direct solver for the least-squares spectral collocation system on rectangular elements for the incompressible Navier-Stokes equations, *Journal of Computational Physics*, 227(9), 2008, pp. 4776–4796.

[18] J.P.KERNEVEZ, *the sentinal method and its application to environmental pollution,* CRC Press, Boca raton, FL, 1997.

[19] A. KIRSCH, *An Introduction to the Mathematical Theory of Inverse Problems*,Springer, New York, NY, USA, 2011.

[20] H.A. LEVINE, the role of crirical exponents in blowup theorems, *SIAM*, 32 ,1990, pp 269–288.

[21] J.L.LIONS, *Sentinelle pour les systèmes distribués à données incomplètes*, Masson, 1992.

[22] S. LUO, J. QIAN and P. STEFANOV. Adjoint state method for the identification problem in SPECT: Recovery of both the source and the attenuation in the attenuated X-ray transform. *SIAM Journal on Imaging Sciences*, 7, 2014, pp. 696–715.

[23] R. MOSE, M.E.STOECKEL, C.POULARD, P.ACKERER et F.LEHMANN, Transport parameters identification: application of the sentinel method, *Computational Geosciences* , 4, 2000, pp. 251–273.

[24] A. TARANTOLA, *Inverse Problem Theory and Methods for Model Parameter Estimation*,SIAM, 2005.

[25] M. A. TAYLOR, B. A. WINGATE, and R. E. VINCENT, An algorithm for computing FEKETE points in the triangle, *SIAM*,38, 2000, pp. 1707–1720.

[26] A. TRAORE, B. MAMPASSI, L. SOME, A least-squares spectral collocation formulation for solving PDEs on complex geometry domains, *International Journal of Applied Mathematics and Computation*,2, 2011, pp. 9–22.

[27] A.TRAORE, B. MAMPASSI, B. SALEY, A Numerical Approach of the sentinel method for distributed parameter systems, *CEJM*, 4; 2007, pp. 751–763.

[28] L.N.TREFETHEN, Spectral Method in Matlab, *Siam*, 2000.

[29] T.C. WARBURTON, S.J.SHERWIN, and G.E. KARNIADAKIS, Basis Functions for Triangular and Quadrilateral High-Order Elements, *SIAM*, 20, 1999, pp. 1671–1695.

[30] J. WU, Least-Squares methods for solving partial differential equations by using Bezier control Points, *Elselvier, Applied mathematics and computer*,219, 2012, pp. 3655–3663.

Step 1 Read the following parameters:

- n : Mesh parameter,
- T : Experience time,
- $\triangle t$: A unit of time,
- N : Degree of interpolation polynomials,
- ε : Precision parameter.

Step 2 Mesh the domains

- Ω : Whole domain,
- Ω_{obs} : Observatory.

Step 3 Compute the vectors and matrices associated to the mesh.

- Z : Assembly matrix,
- D_{bord} : Matrix which select the boundary points, computed according to (24),
- D_{obs} : Matrix which select the observatory points, computed according to (24),
- $\mathbb{L}$: Discrete differential operator approaching the Laplacian such as (12)
- x_{loc} : Vector of local collocation points, obtained according to (2),
- $x_{glob} = Z \times x_{loc}$: Vector of global collocation points,
- $x_{bord} = D_{bord} \times x_{glob}$: Vector of boundary points,
- $x_{obs} = D_{obs} \times x_{glob}$: Vector of observatory points.

Step 4 Define the structure of vectorial functions:

- $\underline{f}(t, \lambda)$ according to (20),
- $\underline{g}(t, \tau)$ according to (21),

Step 5 Define the residual operator

$$\mathscr{L}(\underline{v}, \underline{C}) = \left\| \frac{d\underline{C}(t)}{dt} - \xi \mathbb{L} \times \underline{C}(t) + \eta \underline{C}(t) - \mu |\underline{C}|^p(t) - \underline{f}(t; \lambda) \right\|^2.$$

Step 6 Compute the regularizing parameter $\widehat{\beta}$.

Step 7 Define the least squares spectral collocation operator

$$J_{\widehat{\beta}}(\underline{v}, \underline{C}) = \sum_{k=1}^{[T/\Delta t]} \|D_{obs} \times \underline{C}(k\Delta t) - \underline{C}_{obs}(k\Delta t)\|^2 + \sum_{k=1}^{[T/\Delta t]} \|D_b \times \underline{C}(k\Delta t)\|^2 + \widehat{\beta} \times \|\underline{v} - \widehat{\underline{v}}\|^2 + \mathscr{L}(\underline{v}, \underline{C}).$$

Step 8 Choose an arbitrary value $\underline{v}_0$.

Step 9 Achieve the following problem

$$J_{\widehat{\beta}}(\widehat{\underline{v}}, \widehat{\underline{C}}) = \min_{\underline{v}, \underline{C}} J_{\widehat{\beta}}(\underline{v}, \underline{C}).$$

Figure 5: *Algorithm of least squares spectral collocation scheme.*

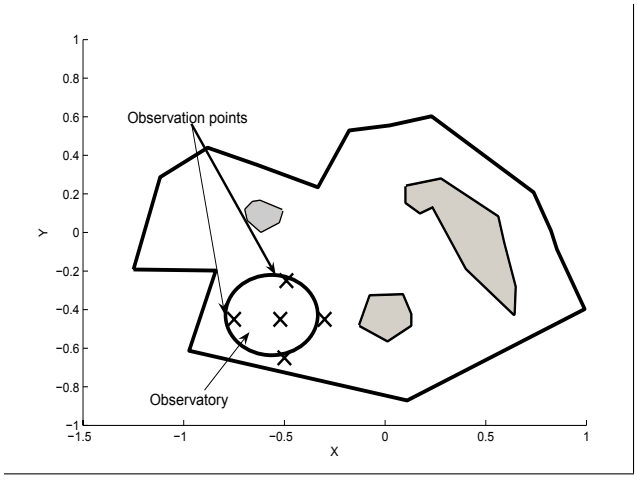


Figure 6: *Presentation of observation domain Ω_{obs} and the points, O_i used for the measurements .*

	Measures			
	O1	O2	O3	O4
t_1	0	0	0	0
t_2	0.0341	1.6315	2.4895	2.4303
t_3	0.1227	0.2481	1.5021	0.4914
t_4	0.2471	0.5347	1.0590	1.6734
t_5	0.4863	0.6602	0.7808	1.2885
t_6	0.8996	2.0906	2.6617	4.3541
t_7	1.9562	3.1453	2.7454	3.3367
t_8	4.7598	4.7221	5.4121	4.7685
t_9	8.0019	8.3292	9.1448	9.5251
t_{10}	10.1644	10.3390	11.0510	10.0418
t_{11}	10.2478	10.0611	11.3513	11.3318
t_{12}	9.9489	10.6747	11.3433	11.2951
t_{13}	11.0906	10.5023	11.6931	11.6932
t_{14}	11.6184	13.8016	12.5992	12.2813
t_{15}	11.0771	10.9407	10.9209	12.0159
t_{16}	12.0918	12.2057	12.0490	10.1750
t_{17}	14.0955	15.1623	13.5582	13.9864
t_{18}	14.1742	14.2334	14.4907	15.3864
t_{19}	15.1680	15.0724	14.0159	14.7469
t_{20}	15.3715	14.5391	15.9543	16.5321
t_{21}	16.5278	16.8222	16.0171	16.0574
t_{22}	16.0663	14.7302	15.2589	15.9360
t_{23}	16.2276	16.9419	17.1612	17.7301
t_{24}	14.0069	15.6304	14.7085	14.4529

Figure 7: *The data obtained at four (4) observation points (see also figure 8).*

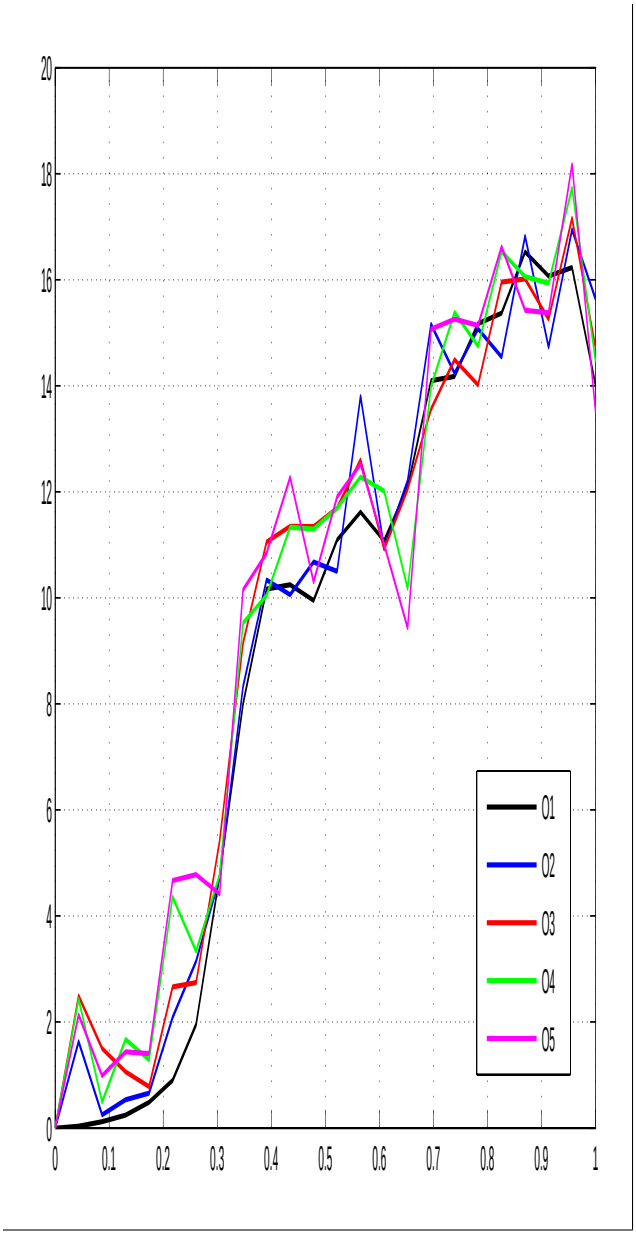


Figure 8: *The measurement curves at 5 observation points* , O_1, O_2, O_3, O_4 and O_5

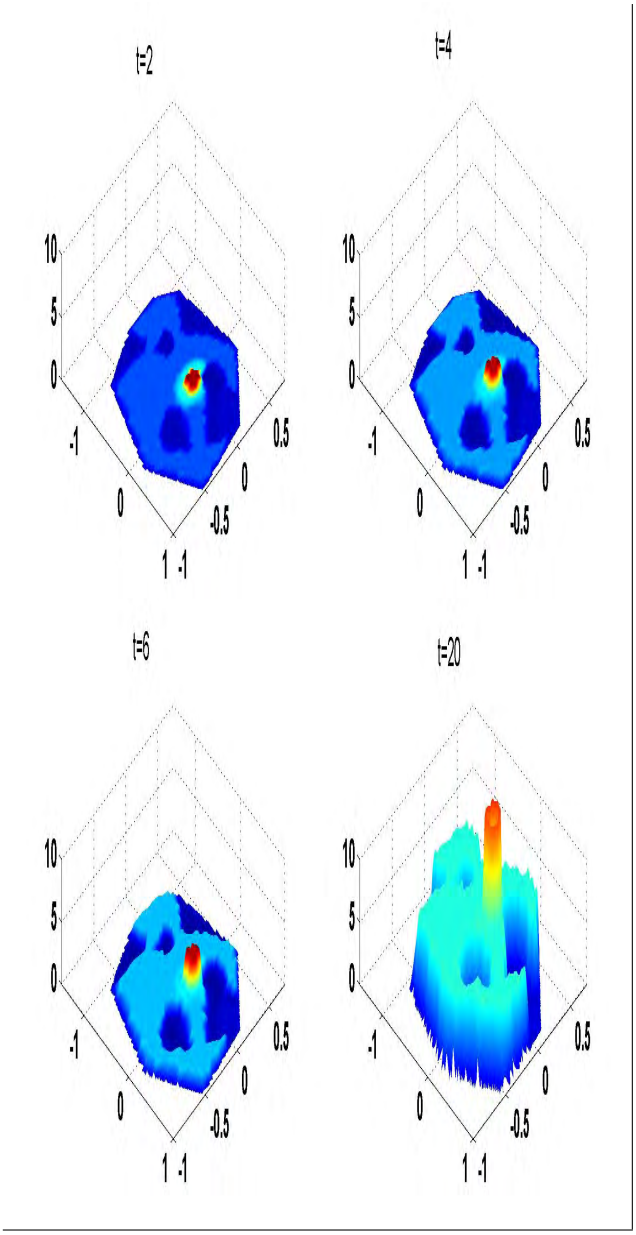


Figure 9: *Concentration curve in* $3D$ *dimension at times* $t = 2$ *respectively* $t = 4$, $t = 6$ *and* $t = 20$.

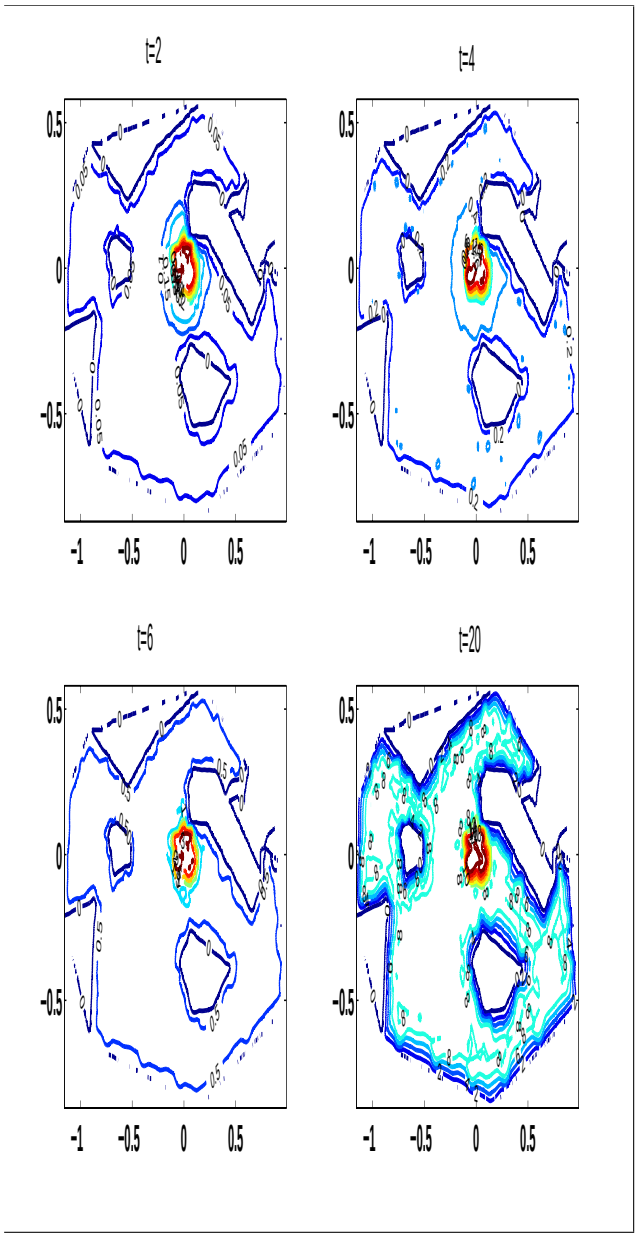


Figure 10: *Presentation of pollution zone at times* $t = 2$ *respectively* $t = 4$, $t = 6$ *and* $t = 20$.

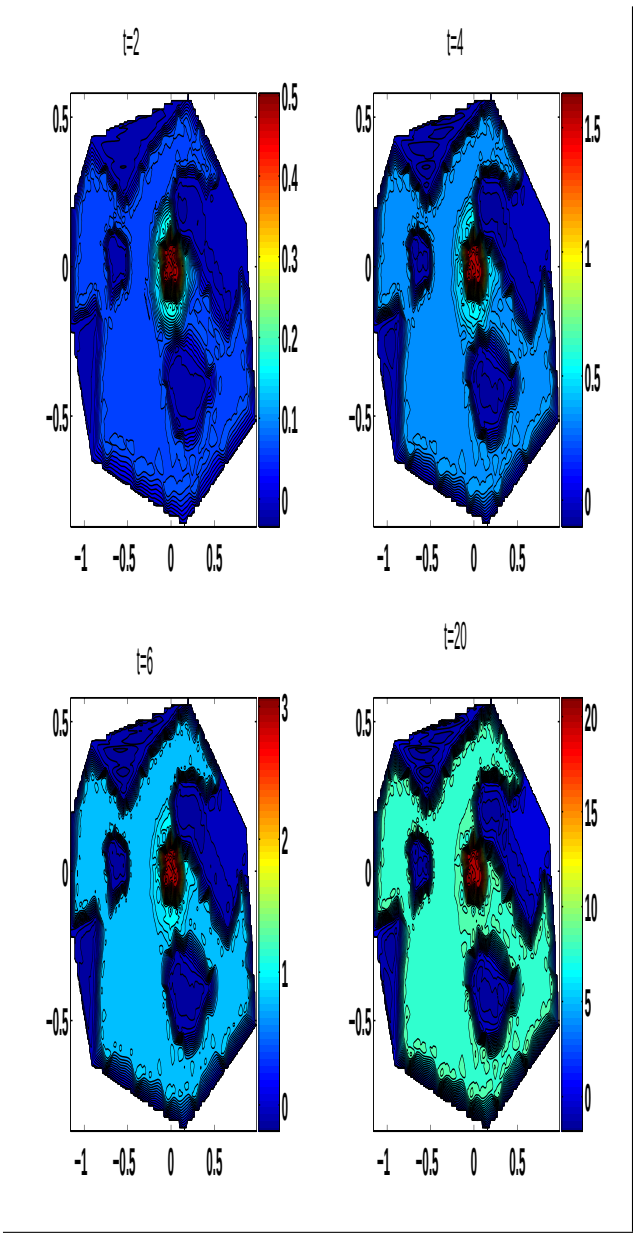


Figure 11: *Presentation of pollution zone at times* $t = 2$ *respectively* $t = 4$, $t = 6$ *and* $t = 20$.

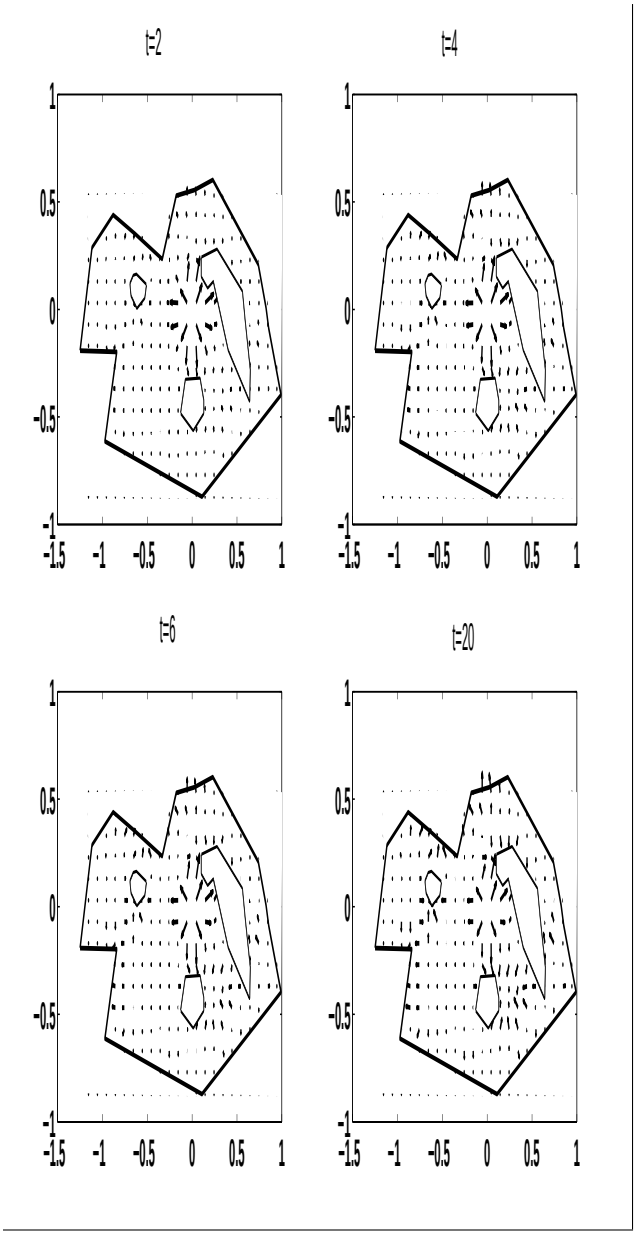


Figure 12: *Presentation of pollution movement sens at times* $t = 2$ *respectively* $t = 4$, $t = 6$ *and* $t = 20$.

ANNEXE 2

Article :

R. Y. M. MASSOUKOU1, A. TRAORE, *An age-structured model for the transmission dynamics of hepatitis B: asymptotic analysis*, Turkish Journal of Mathematics, TÜBITAK, 41,pp 436--460 (2017)

TÜBİTAK

Turkish Journal of Mathematics

http://journals.tubitak.gov.tr/math/

Research Article

Turk J Math
(2017) 41: 436 – 460
© TÜBİTAK
doi:10.3906/mat-1412-50

An age-structured model for the transmission dynamics of hepatitis B: asymptotic analysis

Rodrigue Yves M'PIKA MASSOUKOU[1], **Aboubakari TRAORE**[2,*]
[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa
[2]Department of Mathematics and Computer Sciences, Cheikh Anta Diop University, Dakar, Senegal

Received: 20.12.2014 • **Accepted/Published Online:** 01.06.2016 • **Final Version:** 03.04.2017

Abstract: In this paper, we consider the age-structured model for the transmission dynamics of Hepatitis B virus (HBV) proposed earlier in the article by Zou et al.: An age-structured model for transmission dynamics of hepatitis B. SIAM J Appl Math 2010; 70: 3121-3139, where a slight modification is made. We consider that the HBV infection processes act on a time scale different from that of the vital processes. Such a model becomes a multiple time scale model and thus it often can be significantly simplified by various asymptotic methods. We apply, as in the paper of Banasiak and M'pika Massoukou: A singularly perturbed age structured SIRS model with fast recovery. Disc Cont Dyn Sys 2014, a suitable technique of asymptotic analysis, based on the Chapman–Enskog procedure, which allows separation of scales and aggregation of variables.

Key words: Hepatitis B virus (HBV), age structure, asymptotic analysis

1. Introduction

Hepatitis B virus (HBV) is mainly transmitted through body fluids like blood, semen, and vaginal secretions. One of the most important factors influencing the probability of developing carriage of HBV is age, involving mature individuals. Thus, it can be expected that the interplay of the demographic processes with the infection mechanism will produce a nontrivial dynamics.

In this paper we consider the model for transmission dynamic of HBV introduced in [20], but slightly modified, which is formulated under the following assumptions:

(1) The latent, acute, and carrier stages are differentiated. Only acute individuals and carrier individuals are infectious.

(2) All latently infected individuals develop acute hepatitis B first.

(3) Some individuals with acute infection progress towards the carrier state and later develop immunity while others develop immunity without progressing towards the carrier state.

(4) Since the disease-induced death rate is relatively low, it is ignored.

(5) There is a possibility for treatment (or recovery) during both the acute stage of infection and the carrier state of infection.

*Correspondence: traoreabou08@gmail.com
2010 *AMS Mathematics Subject Classification:* 35Q92; 35B25; 47D03; 92D30.

The population is divided into six subclasses: susceptible individuals, infected individuals but not yet infectious (latent), acutely infectious individuals, carrier individuals, individuals who have recovered from infection and are now immune, and vaccinated immune individuals, and we denote by $s(a,t)$, $l(a,t)$, $i(a,t)$, $c(a,t)$, $r(a,t)$, and $v(a,t)$ the corresponding density functions for these epidemiological age-structured classes, respectively.

Let $\beta(a)$ and $\mu(a)$ be the age-specific fertility and the age-specific mortality (or natural mortality rate) of the population, which we suppose are not affected by the disease; σ is the rate moving from latent to acute, γ_1 is the rate moving from acute to carrier, $\gamma_2(a)$ is the rate moving from carrier to vaccinated, $p(a,t)$ is the vaccination rate against HBV, $(1-\omega)$ is the proportion of births with successful vaccination, $\omega \in [0,1]$, $q(a)$ is the probability an individual fails to clear an acute infection and develops to carrier state, ψ is the rate of waning vaccine-induced immunity, and Λ is the infection rate (or force of infection).

We consider the following separable intercohort constitutive form for the force of infection [7]:

$$\Lambda(a, i(\cdot,t), c(\cdot,t)) = k(a) \int_0^{a_+} [h_1(a) i(a,t) + h_2(a) c(a,t)]\, da, \tag{1}$$

where $h_1(a)$ and $h_2(a)$ are the age-specific infectiousness corresponding to acute stage and chronic stage, respectively, $k(a)$, the age-specific contagion rate; here we assume that the two stages have the same contagion rate and a_+ is the maximum age of an individual. We assume that $h_1(a)$, $h_2(a)$, and $k(a)$ satisfy the following conditions:

$$h_j, k \in \mathrm{L}^\infty([0,a_+]), \quad h_j(a), k(a) \geq 0 \quad \text{a.e. in } [0,a_+]. \tag{2}$$

We assume that there is no proportion of perinatal infected (from carrier mothers) newborns, there is a proportion of births with successful vaccination, there is a proportion of susceptible newborns, and that an individual may become infected through contact both with acute hepatitis B individuals and chronic hepatitis B individuals. Then the dynamics of the age-structured epidemiological model for the transmission of HBV can be described by the following initial boundary value problem:

$$\begin{aligned}
\partial_t s(a,t) &= -\partial_a s(a,t) - \mu(a) s(a,t) + \psi v(a,t) - s(a,t)\Lambda(a, i(\cdot,t), c(\cdot,t)) \\
&\quad - p(a,t) s(a,t), \\
\partial_t l(a,t) &= -\partial_a l(a,t) - \mu(a) l(a,t) - \sigma l(a,t) + s(a,t)\Lambda(a, i(\cdot,t), c(\cdot,t)), \\
\partial_t i(a,t) &= -\partial_a i(a,t) - \mu(a) i(a,t) + \sigma l(a,t) - \gamma_1 i(a,t), \\
\partial_t c(a,t) &= -\partial_a c(a,t) - \mu(a) c(a,t) - \gamma_2(a) c(a,t) + q(a)\gamma_1 i(a,t), \\
\partial_t r(a,t) &= -\partial_a r(a,t) - \mu(a) r(a,t) + \gamma_2(a) c(a,t) + (1-q(a))\gamma_1 i(a,t), \\
\partial_t v(a,t) &= -\partial_a v(a,t) - \mu(a) v(a,t) - \psi v(a,t) + p(a,t) s(a,t),
\end{aligned} \tag{3}$$

with boundary conditions

$$\begin{aligned} s(0,t) &= \omega \int_0^{a_+} \beta(a)\left[s(a,t)+l(a,t)+i(a,t)+r(a,t)+v(a,t)+c(a,t)\right] da, \\ l(0,t) &= i(0,t) = c(0,t) = r(0,t) = 0, \\ v(0,t) &= (1-\omega) \int_0^{a_+} \left[s(a,t)+l(a,t)+i(a,t)+r(a,t)+v(a,t)+c(a,t)\right] da, \end{aligned} \tag{4}$$

and initial conditions

$$\begin{aligned} s(a,0) &= \overset{o}{s}(a), \quad l(a,0) = \overset{o}{l}(a), \quad i(a,0) = \overset{o}{i}(a), \\ c(a,0) &= \overset{o}{c}(a), \quad r(a,0) = \overset{o}{r}(a), \quad v(a,0) = \overset{o}{v}(a). \end{aligned} \tag{5}$$

Note that in building the age-structured (epidemiological) model (3), which takes into account both the vital and infection dynamics, we must be careful as many diseases act on different time scales than the vital process. Since we are dealing with a human population, the death and birth rates are measured in units 1/70 years, where 70 is considered to be the average life-span in the population.

Regarding the numerical values for the parameters in the model (3) provided in [19, 20], we see that $\beta = 0.0121$, $\mu = 0.00693$, $\mu_1 = 0.002$, $\psi = 0.1$, $\Lambda \approx 0.16$, $\sigma = 6/\text{year}$, $\gamma_1 = 4/\text{year}$, and $\gamma_2 = 0.025/\text{year}$. Using 70 years as unit of time in the model (3), the numerical values for σ, γ_1, and γ_2 should be multiplied by 70 and this results in $\sigma = 420$, $\gamma_1 = 280$, and $\gamma_2 = 1.75$. This shows that the processes induced by σ and γ_1 are faster than those induced by β, μ (demography processes), ψ and Λ, while the process induced by γ_2 is slightly faster than those induced by β, μ (demography processes), ψ and Λ. Here we consider the model where the duration of carriage is of the same time-scale as the duration of latency and acute infection such that these processes are faster than those induced by β, μ (demography processes), ψ, and Λ. Thus we consider

$$\begin{aligned} \partial_t \mathbf{u}_\epsilon &= \mathcal{S}\mathbf{u}_\epsilon + \mathcal{M}\mathbf{u}_\epsilon + \mathcal{F}(\mathbf{u}_\epsilon) + \frac{1}{\epsilon}\mathcal{C}\mathbf{u}_\epsilon, \\ \mathbf{u}_\epsilon(0,t) &= \mathcal{B}\left[\mathbf{u}_\epsilon(\cdot,t)\right], \\ \mathbf{u}_\epsilon(a,0) &= \overset{o}{\mathbf{u}}, \end{aligned} \tag{6}$$

where $\mathbf{u}_\epsilon = (s_\epsilon, l_\epsilon, i_\epsilon, c_\epsilon, r_\epsilon, v_\epsilon)$, $\mathcal{S} = \text{diag}\{-\partial_a, -\partial_a, -\partial_a, -\partial_a, -\partial_a, -\partial_a\}$ on $D(\mathcal{S}) = \mathrm{W}^{1,1}\left([0,a_+],\mathbb{R}^6\right)$,

$$\mathcal{M}(a) = \begin{pmatrix} -\mu(a) & 0 & 0 & 0 & 0 & \psi \\ 0 & -\mu(a) & 0 & 0 & 0 & 0 \\ 0 & 0 & -\mu(a) & 0 & 0 & 0 \\ 0 & 0 & 0 & -\mu(a) & 0 & 0 \\ 0 & 0 & 0 & 0 & -\mu(a) & 0 \\ 0 & 0 & 0 & 0 & 0 & -\mu(a)-\psi \end{pmatrix},$$

on $D(\mathcal{M}) = \{\mathbf{u} \in \mathrm{L}^1\left([0,a_+],\mathbb{R}^6\right); \mu\mathbf{u} \in \mathrm{L}^1\left([0,a_+],\mathbb{R}^6\right)\}$ and ϵ is a small parameter reflecting the ratio of the typical time scales of the vital and epidemiological processes. Moreover, the bounded operator $\mathcal{B}$:

$\mathrm{L}^1\left([0,a_+],\mathbb{R}^6\right)\to\mathbb{R}^6$ is defined by

$$\mathcal{B}\mathbf{u}=\int_0^{a_+}B(a)\mathbf{u}(a)\,da,$$

with

$$B(a)=\begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16}\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ b_{61} & b_{62} & b_{63} & b_{64} & b_{65} & b_{66}\end{pmatrix},$$

$b_{1k}=\omega\beta(a)$ and $b_{6k}=(1-\omega)\beta(a)$, $k=1,\dots,6$,

$$\left[\mathcal{F}(\mathbf{u})\right](a)=\begin{pmatrix} -s(a)p(a)-s(a)\Lambda(a,i,c)\\ s(a)\Lambda(a,i,c)\\ 0\\ 0\\ 0\\ s(a)p(a)\end{pmatrix}$$

and

$$\left[\mathcal{C}\mathbf{u}\right](a)=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0\\ 0 & -\sigma & 0 & 0 & 0 & 0\\ 0 & \sigma & -\gamma_1 & 0 & 0 & 0\\ 0 & 0 & q(a)\gamma_1 & -\gamma_2(a) & 0 & 0\\ 0 & 0 & (1-q(a))\gamma_1 & \gamma_2(a) & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}\mathbf{u}(a).$$

We aim to investigate the behavior of the solution $(s_\epsilon,l_\epsilon,i_\epsilon,c_\epsilon,r_\epsilon,v_\epsilon)$, of (6), as $\epsilon\to 0$. We shall note that if $(s_\epsilon,l_\epsilon,i_\epsilon,c_\epsilon,r_\epsilon,v_\epsilon)$ satisfies (6) then, by summing the equations in (6), it results that the total population,

$$n(a,t)=s_\epsilon(a,t)+l_\epsilon(a,t)+i_\epsilon(a,t)+c_\epsilon(a,t)+r_\epsilon(a,t)+v_\epsilon(a,t),$$

satisfies

$$\begin{aligned}\partial_t n(a,t)&=-\partial_a n(a,t)-\mu(a)n(a,t),\\ n(0,t)&=\int_0^{a_+}\beta(a)n(a,t)\,da,\\ n(a,0)&=\overset{\circ}{n}(a),\end{aligned}\tag{7}$$

and is independent of ϵ. It follows from [8] that there exists a dominant eigenvalue $\lambda_\mu\le\overline{\beta}-\underline{\mu}$, defined as the unique real solution to

$$\int_0^{a_+}e^{-\lambda a}\beta(a)\Pi_\mu(a)\,da=1,\tag{8}$$

where, see [2, 8, 16],

$$\Pi_\mu(a) := e^{-\int_0^a \mu(s)\,ds} \tag{9}$$

is the probability of survival of an individual until age a, and a constant M such that

$$\|n(t)\| \leq Me^{\lambda_\mu t}\|\overset{\circ}{n}\|, \tag{10}$$

where λ_μ is negative, zero, or positive if and only if the *net reproduction rate* $R_\mu = \int_0^{a_+} \beta(a)\Pi_\mu(a)\,da$ is, respectively, smaller than, equal to, or greater than one. As we mentioned earlier, the considered scaling is realistic for models in which there are no significant changes in the total population. Hence, throughout the paper we will assume

$$R_\mu \leq 1, \tag{11}$$

that is, $\overline{\beta} \leq \underline{\mu}$.

Let $\overline{w}$ be the solution to the McKendrick–von Foerster problem

$$\begin{aligned} \partial_t \overline{w}(a,t) &= -\partial_a \overline{w}(a,t) - \mu(a)\overline{w}(a,t), \\ \overline{w}(0,t) &= 0, \\ \overline{w}(a,0) &= \overset{\circ}{w}(a) = \overset{\circ}{l}(a) + \overset{\circ}{i}(a) + \overset{\circ}{c}(a) + \overset{\circ}{r}(a), \end{aligned} \tag{12}$$

and $\overline{v}$ be the solution to the McKendrick–von Foerster problem

$$\begin{aligned} \partial_t \overline{v}(a,t) &= -\partial_a \overline{v}(a,t) - \mu(a)\overline{v}(a,t) - \psi\overline{v}(a,t) \\ &\quad - p(a,t)\overline{v}(a,t) + p(a,t)(n(a,t) - \overline{w}(a,t)), \\ \overline{v}(0,t) &= (1-\omega)n(0,t), \\ \overline{v}(a,0) &= \overset{\circ}{v}(a). \end{aligned} \tag{13}$$

2. Notation, assumptions, and well-posedness results

In the sequel we consider the state space $\mathbf{X} = \mathrm{L}^1\left([0,a_+],\mathbb{R}^6\right)$. We denote by $\mathbf{X}_+ = \mathrm{L}^1\left([0,a_+],\mathbb{R}^6_+\right)$ the positive cone of $\mathbf{X}$ and $\|\cdot\|$ the suitable norm in L^1. If necessary, the notation $\|\cdot\|_X$ will be used for the norm in a specific space X. In addition, for any measurable function η on $[0,a_+]$, we introduce the notation

$$\overline{\eta} = \operatorname*{ess\,sup}_{a\in[0,a_+]} \eta(a), \quad \underline{\eta} = \operatorname*{ess\,inf}_{a\in[0,a_+]} \eta(a)$$

and we make the assumptions

A1: $\mu \in \mathrm{L}^1_{\mathrm{loc}}\left([0,a_+)\right)$, $\int_0^{a_+} \mu(s)\,ds = \infty$ with $\underline{\mu} > 0$;

A2: $\beta \in \mathrm{L}^\infty\left([0,a_+]\right)$;

A3: $q, \gamma_2 \in \mathrm{W}^{1,\infty}([0,a_+])$ with $\underline{q} > 0$, $\underline{\gamma_2} > 0$;

A4: $p \in \mathcal{C}([0,a_+] \times [0,T])$.

We make the biologically realistic assumption on the maximum age, a_+, such that $a_+ < \infty$, meaning that no individual can live indefinitely. This requires the survival probability Π_μ, defined by

$$\Pi_\mu(a) := e^{-\int_0^a \mu(s)\,ds}, \tag{14}$$

to satisfy $\Pi_\mu(a_+) = 0$, which accounts for the nonintegrable singularity in **A1**. Hence, μ cannot be assumed bounded as $a \to a_+^-$ while it could be bounded in the case $a_+ = \infty$. This justifies the fact that most authors [9, 11, 14, 15, 18] consider an infinite maximum age, $a_+ = \infty$, though without any biological significance, in order to be able to handle μ after being assumed bounded. In some papers, e.g., [10], the unboundedness of μ, in the case $a_+ < \infty$, was circumvented by assuming that there is a maximum reproductive age $a_r < a_+$, so that the birth rate satisfies $\beta(a) = 0$ for $a > a_r$, and hence ignoring the postreproductive population by performing the analysis for $a \in [0, a_r]$, which causes the loss of the conservativeness of the model. The analysis of the model without any simplifying assumption in the scalar linear case was done in [8] by reducing it to an integral equation along the characteristics. It is known that the solutions obtained in this way generate a strongly continuous semigroup on $\mathrm{L}^1([0,a_+])$; see [3, 17] (though in [17] it is assumed that $a_+ = \infty$.)

The technical details required to handle the unbounded μ, on $[0,a_+]$ with $a_+ < \infty$, in the problem at hand can be found in [12]. In particular, it follows that the realization of the operator $\mathcal{A} := \mathcal{S} + \mathcal{M}$ on the domain

$$D(\mathcal{A}) = \{\mathbf{u} \in D(\mathcal{S}) \cap D(\mathcal{M});\ \mathbf{u}(0) = \mathcal{B}\mathbf{u}\} \tag{15}$$

generates a positive $\mathcal{C}_0$-semigroup, denoted by $\left(e^{t\mathcal{A}}\right)_{t\geq 0}$. Since, for a fixed ϵ, (7) is a quadratic perturbation of the linear system, a standard argument, see [5, 12], shows that if $\overset{\circ}{\mathbf{u}} = (\overset{\circ}{s}, \overset{\circ}{l}, \overset{\circ}{i}, \overset{\circ}{c}, \overset{\circ}{r}, \overset{\circ}{v}) \in \mathbf{X}_+$, then there exists a unique global positive mild solution $t \to \mathbf{u}_\epsilon(t) = (s_\epsilon(t), l_\epsilon(t), i_\epsilon(t), c_\epsilon(t), r_\epsilon(t), v_\epsilon(t)) \in \mathcal{C}([0,\infty), \mathbf{X})$ to (7). This solution becomes a classical solution if $\overset{\circ}{\mathbf{u}} \in D(\mathcal{A})$. In such a case we obtain, in particular, that $\mathbf{u}_\epsilon$ is continuous on $[0,a_+] \times [0,T]$ for any $0 \leq T < \infty$, $\mathbf{u}_\epsilon \in D(\mathcal{S}) \cap D(\mathcal{M})$ and (7) is satisfied termwise almost everywhere on $[0,a_+] \times [0,T]$. Standard calculations, see e.g. [10], show that $\left(e^{t\mathcal{A}}\right)_{t\geq 0}$ satisfies the estimate

$$\|e^{t\mathcal{A}}\| \leq e^{(\overline{\beta} - \underline{\mu})t}. \tag{16}$$

This estimate can be improved. In fact, as mentioned earlier, if $(s_\epsilon(t), l_\epsilon(t), i_\epsilon(t), c_\epsilon(t), r_\epsilon(t), v_\epsilon(t))$ is a classical solution to (6), then $n(a,t) = s_\epsilon(a,t) + l_\epsilon(a,t) + i_\epsilon(a,t) + c_\epsilon(a,t) + r_\epsilon(a,t) + v_\epsilon(a,t)$ is a classical solution to (7). We denote by $(A, D(A))$ the generator of the semigroup $\left(e^{tA}\right)_{t\geq 0}$ for (7), with domain defined analogously to (15). Hence, since $\overset{\circ}{\mathbf{u}} \geq 0$ yields $\mathbf{u}_\epsilon(t) = (s_\epsilon(t), l_\epsilon(t), i_\epsilon(t), c_\epsilon(t), r_\epsilon(t), v_\epsilon(t)) \geq 0$, we see that each component of $\mathbf{u}_\epsilon$ is controlled by the ϵ-independent solution n of (7):

$$\begin{array}{lll} 0 \leq s_\epsilon(a,t) \leq n(a,t), & 0 \leq l_\epsilon(a,t) \leq n(a,t), & 0 \leq i_\epsilon(a,t) \leq n(a,t), \\ 0 \leq c_\epsilon(a,t) \leq n(a,t), & 0 \leq r_\epsilon(a,t) \leq n(a,t) \quad \text{and} & 0 \leq v_\epsilon(a,t) \leq n(a,t) \end{array} \tag{17}$$

for each $t \geq 0$ and almost every $a \in [0,a_+]$ thus satisfying the inequality (10).

3. Formal asymptotic expansion

Following the general approach of the asymptotic analysis, see e.g. [4, 6], we are looking for the so-called *hydrodynamic space* V of the singularly perturbed equation (6), which, in this case, is given by the null-space of $\mathcal{C}$. Since, in this context, a is treated as a parameter, we perform the calculations for a fixed a. Then we have

$$V = \{\mathbf{u} \in \mathbb{R}^6;\ \mathbf{u} = (u_1, 0, 0, 0, u_5, u_6), u_1, u_5, u_6 \in \mathbb{R}\} = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\},$$

where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are an approximate basis for V. The complementary spectral space W, called the *kinetic space*, corresponding to the eigenvalues $\lambda_4 = -\sigma$, $\lambda_5 = -\gamma_1$, and $\lambda_6 = -\gamma_2(a)$, respectively, is spanned by $\mathbf{e}_4$, $\mathbf{e}_5$, and $\mathbf{e}_6$ given by

$$\begin{pmatrix} 0 \\ 1 \\ -\frac{\sigma}{\sigma-\gamma_1} \\ \frac{\sigma q(a)\gamma_1}{(\sigma-\gamma_1)(\sigma-\gamma_2(a))} \\ \frac{\gamma_1}{\sigma-\gamma_1}\cdot\frac{(1-q(a))\sigma-\gamma_2(a)}{\sigma-\gamma_2(a)} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -\frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)} \\ -\frac{(1-q(a))\gamma_1-\gamma_2(a)}{\gamma_1-\gamma_2(a)} \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix},$$

respectively. To come up with the decomposition $\mathbf{u} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3 + c_4\mathbf{e}_4 + c_5\mathbf{e}_5 + c_6\mathbf{e}_6$ in the hydrodynamic and kinetic space, we have to find the coefficients c_i, $i = 1, \ldots, 6$, which yield the aggregated variables. Using standard results from linear algebra, see e.g. [6], $c_i = \mathbf{f}_i \cdot \mathbf{u}/\mathbf{f}_i \cdot \mathbf{e}_i$, where $\mathbf{f}_i$ are left eigenvectors of $\mathcal{C}$ corresponding to the eigenvectors as $\mathbf{e}_i$, $i = 1, \ldots, 6$. Here the context is more complicated as V is three dimensional and we have to choose its basis in an appropriate way. Since $\mathbf{f}_1 = (1, 1, 1, 1, 1, 1)$ is a left eigenvector of $\mathcal{C}$ corresponding to the zero eigenvalue and $\mathbf{f}_1 \cdot \mathbf{u} = u_1 + u_2 + u_3 + u_4 + u_5 + u_6$, it should play a significant role in the analysis due to (7). Then, for the convenience of calculations, we take $\mathbf{e}_2 = (-1, 0, 0, 0, 0, 1) \in V$, which is orthogonal to $\mathbf{f}_1$. We have some freedom in selecting $\mathbf{f}_3$: taking $\mathbf{f}_3 = (0, 0, 0, 0, 0, 1)$ yields $\mathbf{e}_1 = (1, 0, 0, 0, 0, 0) \in V$ and $\mathbf{e}_2 = (-1, 0, 0, 0, 1, 0) \in V$. Finally $\mathbf{f}_4 = (0, 1, 0, 0, 0, 0)$, $\mathbf{f}_5 = (0, \sigma/(\sigma-\gamma_1), 1, 0, 0, 0)$, $\mathbf{f}_6 = (0, \sigma q(a)\gamma_1/(\sigma-\gamma_2(a))(\gamma_1-\gamma_2(a)), q(a)\gamma_1/(\gamma_1-\gamma_2(a)), 1, 0, 0)$ so that

$$\begin{aligned} \mathbf{u} = {} & (u_1+u_2+u_3+u_4+u_5+u_6)\mathbf{e}_1 + (u_2+u_3+u_4+u_5)\mathbf{e}_2 + u_6\mathbf{e}_3 + u_2\mathbf{e}_4 \\ & + \left(\frac{\sigma}{\sigma-\gamma_1}u_2 + u_3\right)\mathbf{e}_5 + \left(\frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\sigma}{\sigma-\gamma_1}u_2 + \frac{q\gamma_1}{\gamma_1-\gamma_2}u_3 + u_4\right)\mathbf{e}_6. \end{aligned} \tag{18}$$

Then we use this decomposition to change variables in (6). Accordingly, we define $n = s_\epsilon + l_\epsilon + i_\epsilon + c_\epsilon + r_\epsilon + v_\epsilon$, $w_\epsilon = l_\epsilon + i_\epsilon + c_\epsilon + r_\epsilon$, $z_\epsilon = \frac{\sigma}{\sigma-\gamma_1}l_\epsilon + i_\epsilon$, $x_\epsilon = \frac{q\gamma_1}{\gamma_1-\gamma_2}z_\epsilon + c_\epsilon$ and leave l_ϵ, v_ϵ unchanged. This yields the system

of equations

$$\begin{aligned}
\partial_t n(a,t) &= -\partial_a n(a,t) - \mu(a)n(a,t),\\
\partial_t w_\epsilon(a,t) &= -\partial_a w_\epsilon(a,t) - \mu(a)w_\epsilon(a,t) + \Lambda(a, l_\epsilon(\cdot,t), x_\epsilon(\cdot,t), z_\epsilon(\cdot,t))\\
&\quad \times (n(a,t) - w_\epsilon(a,t) - v_\epsilon(a,t)),\\
\partial_t z_\epsilon(a,t) &= -\partial_a z_\epsilon(a,t) - \mu(a)z_\epsilon(a,t) - \frac{\gamma_1}{\epsilon} z_\epsilon(a,t)\\
&\quad + \frac{\sigma}{\sigma-\gamma_1}\Lambda(a, l_\epsilon(\cdot,t), x_\epsilon(\cdot,t), z_\epsilon(\cdot,t))(n(a,t) - w_\epsilon(a,t) - v_\epsilon(a,t)),\\
\partial_t x_\epsilon(a,t) &= -\partial_a x_\epsilon(a,t) - \mu(a)x_\epsilon(a,t) - \frac{\gamma_2(a)}{\epsilon} x_\epsilon(a,t) - \frac{\gamma_1}{\epsilon}\cdot\frac{\sigma q(a)}{\sigma-\gamma_1} l_\epsilon(a,t)\\
&\quad + \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\cdot\frac{\gamma_2'(a)}{\gamma_1-\gamma_2(a)} z_\epsilon(a,t) + \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\\
&\quad \times \Lambda(a, l_\epsilon(\cdot,t), x_\epsilon(\cdot,t), z_\epsilon(\cdot,t))(n(a,t) - w_\epsilon(a,t) - v_\epsilon(a,t)),\\
\partial_t l_\epsilon(a,t) &= -\partial_a l_\epsilon(a,t) - \mu(a)l_\epsilon(a,t) - \frac{\sigma}{\epsilon} l_\epsilon(a,t)\\
&\quad + \Lambda(a, l_\epsilon(\cdot,t), x_\epsilon(\cdot,t), z_\epsilon(\cdot,t))(n(a,t) - w_\epsilon(a,t) - v_\epsilon(a,t)),\\
\partial_t v_\epsilon(a,t) &= -\partial_a v_\epsilon(a,t) - \mu(a)v_\epsilon(a,t) - \psi v_\epsilon(a,t) - p(a,t)v_\epsilon(a,t)\\
&\quad + p(a,t)(n(a,t) - w_\epsilon(a,t)),
\end{aligned} \tag{19}$$

$$\begin{aligned}
n(0,t) &= \int_0^{a_+} \beta(a)n(a,t)\,da,\\
w_\epsilon(0,t) &= x_\epsilon(0,t) = z_\epsilon(0,t) = l_\epsilon(0,t) = 0,\\
v_\epsilon(0,t) &= (1-\omega)n(0,t),
\end{aligned} \tag{20}$$

and

$$\begin{aligned}
n(a,0) &= \overset{\circ}{n}(a) = \overset{\circ}{s}(a) + \overset{\circ}{l}(a) + \overset{\circ}{i}(a) + \overset{\circ}{c}(a) + \overset{\circ}{r}(a) + \overset{\circ}{v}(a),\\
w_\epsilon(a,0) &= \overset{\circ}{w}(a) = \overset{\circ}{l}(a) + \overset{\circ}{i}(a) + \overset{\circ}{c}(a) + \overset{\circ}{r}(a),\\
z_\epsilon(a,0) &= \overset{\circ}{z}(a) = \frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \overset{\circ}{i}(a),\\
x_\epsilon(a,0) &= \overset{\circ}{x}(a) = \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\cdot\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\overset{\circ}{i}(a) + \overset{\circ}{c}(a),\\
l_\epsilon(a,0) &= \overset{\circ}{l}(a), \quad v_\epsilon(a,0) = \overset{\circ}{v}(a),
\end{aligned} \tag{21}$$

where

$$\Lambda(a,l_\epsilon(\cdot,t),x_\epsilon(\cdot,t),z_\epsilon(\cdot,t)) = k(a)\int_0^{a_+}\left[-\frac{\sigma}{\sigma-\gamma_1}h_1(a)l_\epsilon(a,t)+h_2(a)x_\epsilon(a,t)\right.$$
$$\left.+\left(h_1(a)-\frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}h_2(a)\right)z_\epsilon(a,t)\right]da.$$

Therefore, the triplet $(n,w_\epsilon,v_\epsilon)\in V$ and the triplet $(z_\epsilon,x_\epsilon,l_\epsilon)\in W$. We see that the total population n decouples from the system, it is unnecessary to approximate it, and it can be treated as a known function. Therefore we shall focus on the quintuplet $(w_\epsilon,z_\epsilon,x_\epsilon,l_\epsilon,v_\epsilon)$. Let $(\overline{w},\overline{z},\overline{x},\overline{l},\overline{v})$ be the bulk approximation of the quintuplet $(w_\epsilon,z_\epsilon,x_\epsilon,l_\epsilon,v_\epsilon)$, we note that $(w_\epsilon,z_\epsilon,x_\epsilon,l_\epsilon,v_\epsilon)\approx(\overline{w},\overline{z},\overline{x},\overline{l},\overline{v})$, where $\overline{w}$ and $\overline{v}$ are defined by (12) and (13), respectively, and the approximate equality symbol $\approx$ accounts for the fact that we only consider the first terms of the asymptotic expansion. Following the Chapman–Enskog procedure, we only expand the kinetic part of the bulk approximation such that

$$\overline{z}=\overline{z}_0+\epsilon\overline{z}_1+\cdots,\ \overline{x}=\overline{x}_0+\epsilon\overline{x}_1+\cdots,\ \overline{l}=\overline{l}_0+\epsilon\overline{l}_1+\cdots. \tag{22}$$

Inserting (22) into the last four equations in (19), we get

$$\partial_t\overline{z}_0+\epsilon\partial_t\overline{z}_1=-\partial_a\overline{z}_0-\epsilon\partial_a\overline{z}_1-\mu\overline{z}_0-\epsilon\mu\overline{z}_1-\frac{\gamma_1}{\epsilon}\overline{z}_0-\gamma_1\overline{z}_1+\frac{\sigma}{\sigma-\gamma_1}$$
$$\times\left(\Lambda(\overline{l}_0,\overline{x}_0,\overline{z}_0)+\epsilon\Lambda(\overline{l}_1,\overline{x}_1,\overline{z}_1)\right)(n-\overline{w}-\overline{v})+O(\epsilon^2),$$
$$\partial_t\overline{x}_0+\epsilon\partial_t\overline{x}_1=-\partial_a\overline{x}_0-\epsilon\partial_a\overline{x}_1-\mu\overline{x}_0-\epsilon\mu\overline{x}_1-\frac{\gamma_2}{\epsilon}\overline{x}_0-\gamma_2\overline{x}_1$$
$$+\frac{1}{\epsilon}\cdot\frac{q\gamma_1^2}{\gamma_1-\gamma_2}\overline{z}_0+\frac{q\gamma_1^2}{\gamma_1-\gamma_2}\overline{z}_1-\frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{q\gamma_2^{'}}{\gamma_1-\gamma_2}\overline{z}_0$$
$$-\epsilon\frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{q\gamma_2^{'}}{\gamma_1-\gamma_2}\overline{z}_1-\frac{q}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma-\gamma_1}\overline{l}_0-\frac{\sigma q\gamma_1}{\sigma-\gamma_1}\overline{l}_1$$
$$+\frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\sigma}{\sigma-\gamma_1}\left(\Lambda(\overline{l}_0,\overline{x}_0,\overline{z}_0)+\epsilon\Lambda(\overline{l}_1,\overline{x}_1,\overline{z}_1)\right)$$
$$\times(n-\overline{w}-\overline{v})+O(\epsilon^2),$$
$$\partial_t\overline{l}_0+\epsilon\partial_t\overline{l}_1=-\partial_a\overline{l}_0-\epsilon\partial_a\overline{l}_1-\mu\overline{l}_0-\epsilon\mu\overline{l}_1+\left(\Lambda(\overline{l}_0,\overline{x}_0,\overline{z}_0)+\epsilon\Lambda(\overline{l}_1,\overline{x}_1,\overline{z}_1)\right)$$
$$\times(n-\overline{w}-\overline{v})-\frac{\sigma}{\epsilon}\overline{l}_0-\sigma\overline{l}_1+O(\epsilon^2).$$

Comparing coefficients at like powers of ϵ and using $\Lambda(0,0,0)=0$ we get $\overline{z}_0=\overline{x}_0=\overline{l}_0=0$, $\overline{z}_1=\overline{x}_1=\overline{l}_1=0$.

Hence we arrive at the (formal) bulk approximation

$$(n,\overline{w},\overline{z},\overline{x},\overline{l},\overline{v})=(n,\overline{w},0,0,0,\overline{v}).$$

Note that by substituting $l_\epsilon\approx\overline{l}=0$, $x_\epsilon\approx\overline{x}=0$, $z_\epsilon\approx\overline{z}=0$ in the second equation in (19) we arrive at the system, (12). Furthermore, substituting $\overline{w}$ into the last equation in (19) leads to (13). The error of the approximation,

$$(n,w_\epsilon,z_\epsilon,x_\epsilon,l_\epsilon,v_\epsilon)\approx(n,\overline{w},0,0,0,\overline{v}),$$

denoted by

$$\begin{aligned}\overline{\mathbf{E}} = (\overline{e}_w, \overline{e}_z, \overline{e}_x, \overline{e}_l, \overline{e}_v) &= (w_\epsilon - \overline{w}, z_\epsilon - \overline{z}, x_\epsilon - \overline{x}, l_\epsilon - \overline{l}, v_\epsilon - \overline{v}) \\ &= (w_\epsilon - \overline{w}, z_\epsilon, x_\epsilon, l_\epsilon, v_\epsilon - \overline{v}),\end{aligned} \tag{23}$$

satisfies

$$\begin{aligned}
\partial_t \overline{e}_w &= -\partial_a \overline{e}_w - \mu \overline{e}_w - \Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(\overline{e}_w + \overline{e}_v\right) + \Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(n - \overline{w} - \overline{v}\right), \\
\partial_t \overline{e}_z &= -\partial_a \overline{e}_z - \mu \overline{e}_z - \frac{\gamma_1}{\epsilon}\overline{e}_z - \frac{\sigma}{\sigma - \gamma_1}\Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(\overline{e}_w + \overline{e}_v\right) \\
&\quad + \frac{\sigma}{\sigma - \gamma_1}\Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(n - \overline{w} - \overline{v}\right), \\
\partial_t \overline{e}_x &= -\partial_a \overline{e}_x - \mu \overline{e}_x - \frac{\gamma_2}{\epsilon}\overline{e}_x - \frac{1}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma - \gamma_1}\overline{e}_l + \frac{q\gamma_1}{\gamma_1 - \gamma_2}\cdot\frac{\gamma_2^{'}}{\gamma_1 - \gamma_2}\overline{e}_z \\
&\quad - \frac{\sigma}{\sigma - \gamma_1}\cdot\frac{q\gamma_1}{\gamma_1 - \gamma_2}\Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(\overline{e}_w + \overline{e}_v\right) \\
&\quad + \frac{\sigma}{\sigma - \gamma_1}\cdot\frac{q\gamma_1}{\gamma_1 - \gamma_2}\Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(n - \overline{w} - \overline{v}\right), \\
\partial_t \overline{e}_l &= -\partial_a \overline{e}_l - \mu \overline{e}_l - \frac{\sigma}{\epsilon}\overline{e}_l - \Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(\overline{e}_w + \overline{e}_v\right) \\
&\quad + \Lambda\left(\overline{e}_l, \overline{e}_x, \overline{e}_z\right)\left(n - \overline{w} - \overline{v}\right), \\
\partial_t \overline{e}_v &= -\partial_a \overline{e}_v - \mu \overline{e}_v - (p + \psi)\overline{e}_v - p\overline{e}_w,
\end{aligned} \tag{24}$$

with the boundary condition

$$\overline{e}_w(0,t) = \overline{e}_z(0,t) = \overline{e}_x(0,t) = \overline{e}_l(0,t) = \overline{e}_v(0,t) = 0, \tag{25}$$

and the initial condition

$$\begin{aligned}
&\overline{e}_w(a,0) = 0, \quad \overline{e}_z(a,0) = \frac{\sigma}{\sigma - \gamma_1}\overset{\circ}{l}(a) + \overset{\circ}{i}(a), \\
&\overline{e}_x(a,0) = \frac{q\gamma_1}{\gamma_1 - \gamma_2(a)}\cdot\frac{\sigma}{\sigma - \gamma_1}\overset{\circ}{l}(a) + \frac{q\gamma_1}{\gamma_1 - \gamma_2(a)}\overset{\circ}{i}(a) + \overset{\circ}{c}(a), \\
&\overline{e}_l(a,0) = \overset{\circ}{l}(a), \quad \overline{e}_v(a,0) = 0.
\end{aligned} \tag{26}$$

We see that the initial condition is of order 1; therefore the error cannot be of order ϵ. To remedy the situation we have to introduce layer corrections that will take care of the transient phenomena occurring close to $t = 0$, namely the initial layer.

We carry out the initial layer correction by blowing up time according to $\tau = t/\epsilon$ and looking for the approximation

$$(w_\epsilon(t), z_\epsilon(t), x_\epsilon(t), l_\epsilon(t), v_\epsilon(t)) \approx (\overline{w}(t), \widetilde{z}(\tau), \widetilde{x}(\tau), \widetilde{l}(\tau), \overline{v}(t)),$$

where we anticipate that it is unnecessary to introduce the initial layer for w_ϵ and v_ϵ as $\overline{w}$ and $\overline{v}$ satisfy the exact initial condition. We insert the formal expansion

$$\widetilde{z} = \widetilde{z}_0 + \epsilon\widetilde{z}_1 + \cdots, \ \widetilde{x} = \widetilde{x}_0 + \epsilon\widetilde{x}_1 + \cdots, \ \widetilde{l} = \widetilde{l}_0 + \epsilon\widetilde{l}_1 + \cdots$$

and rescale time, due to $\partial_t = \epsilon^{-1}\partial_\tau$, in the third, fourth, and fifth equations in (19).

Comparing coefficients at like powers of ϵ, the equations for the terms at ϵ^{-1} level are

$$\partial_\tau \widetilde{z}_0 = -\gamma_1 \widetilde{z}_0, \quad \partial_\tau \widetilde{l}_0 = -\sigma \widetilde{l}_0,$$
$$\partial_\tau \widetilde{x}_0 = -\gamma_2(a)\widetilde{x}_0 - \frac{\sigma\gamma_1}{\sigma-\gamma_1}\widetilde{l}_0$$

which, subject to the initial condition

$$\widetilde{z}_0(0) = \frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l} + \overset{\circ}{i}, \quad \widetilde{l}_0(0) = \overset{\circ}{l},$$
$$\widetilde{x}_0(0) = \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\cdot\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l} + \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\overset{\circ}{i} + \overset{\circ}{c},$$

yield

$$\begin{aligned}
\widetilde{z}_0(a,t/\epsilon) &= \left(\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \overset{\circ}{i}(a)\right)e^{-\frac{\gamma_1}{\epsilon}t}, \quad \widetilde{l}_0(a,t/\epsilon) = \overset{\circ}{l}(a)e^{-\frac{\sigma}{\epsilon}t}, \\
\widetilde{x}_0(a,t/\epsilon) &= \left(\frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\cdot\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(a) + \frac{q(a)\gamma_1}{\gamma_1-\gamma_2(a)}\overset{\circ}{i}(a) + \overset{\circ}{c}(a)\right)e^{-\frac{\gamma_2(a)}{\epsilon}t} \\
&\quad + \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{\overset{\circ}{l}(a)}{\sigma-\gamma_2(a)}\left(e^{-\frac{\sigma}{\epsilon}t} - e^{-\frac{\gamma_2(a)}{\epsilon}t}\right).
\end{aligned} \tag{27}$$

The new error is given by

$$\begin{aligned}
\widetilde{\mathbf{E}} = (\widetilde{e}_w, \widetilde{e}_z, \widetilde{e}_x, \widetilde{e}_l, \widetilde{e}_v) &= (w_\epsilon - \overline{w}, z_\epsilon - \widetilde{z}_0, x_\epsilon - \widetilde{x}_0, l_\epsilon - \widetilde{l}_0, v_\epsilon - \overline{v}) \\
&= (\overline{e}_w, \overline{e}_z - \widetilde{z}_0, \overline{e}_x - \widetilde{x}_0, \overline{e}_l - \widetilde{l}_0, \overline{e}_v).
\end{aligned} \tag{28}$$

Since $\overset{\circ}{l}, \overset{\circ}{i}, \overset{\circ}{c}, \overset{\circ}{v} \in \mathrm{W}^{1,1}([0,a_+])$ and $\mu\overset{\circ}{l}, \mu\overset{\circ}{i}, \mu\overset{\circ}{c}, \mu\overset{\circ}{v} \in \mathrm{L}^1([0,a_+])$, the error equation for $\widetilde{E}$ can be obtained from (24), (25), and (26) by expressing $\overline{e}_w$, $\overline{e}_z$, $\overline{e}_x$, $\overline{e}_l$, and $\overline{e}_v$ in terms of $\widetilde{e}_w$, $\widetilde{e}_z$, $\widetilde{e}_x$, $\widetilde{e}_l$, and $\widetilde{e}_v$, according to (28). We get

$$\begin{aligned}
\partial_t \widetilde{e}_w &= -\partial_a \widetilde{e}_w - \mu\widetilde{e}_w - \Lambda(\widetilde{e}_l, \widetilde{e}_x, \widetilde{e}_z)(\widetilde{e}_w + \widetilde{e}_v) - \Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)(\widetilde{e}_w + \widetilde{e}_v) \\
&\quad + \Lambda(\widetilde{e}_l, \widetilde{e}_x, \widetilde{e}_z)(n - \overline{w} - \overline{v}) - \Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)(n - \overline{w} - \overline{v}), \\
\partial_t \widetilde{e}_z &= -\partial_a \widetilde{e}_z - \mu\widetilde{e}_z - \frac{\gamma_1}{\epsilon}\widetilde{e}_z - \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{e}_l, \widetilde{e}_x, \widetilde{e}_z)(\widetilde{e}_w + \widetilde{e}_v) \\
&\quad - \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)(\widetilde{e}_w + \widetilde{e}_v) + \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{e}_l, \widetilde{e}_x, \widetilde{e}_z)(n - \overline{w} - \overline{v}) \\
&\quad - \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0)(n - \overline{w} - \overline{v}) - \partial_a \widetilde{z}_0 - \mu\widetilde{z}_0,
\end{aligned}$$

$$
\begin{aligned}
\partial_t \widetilde{e}_x &= -\partial_a \widetilde{e}_x - \mu \widetilde{e}_x - \frac{\gamma_2}{\epsilon}\widetilde{e}_x - \frac{1}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma-\gamma_1}\widetilde{e}_l + \frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\gamma_2^{'}}{\gamma_1-\gamma_2}\widetilde{e}_z \\
&\quad - \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}\Big(\Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)+\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\Big)(\widetilde{e}_w+\widetilde{e}_v) \\
&\quad + \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}\Big(\Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)-\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\Big)(n-\overline{w}-\overline{v}) \\
&\quad + \frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\gamma_2^{'}}{\gamma_1-\gamma_2}\widetilde{z}_0 - \partial_a\widetilde{x}_0 - \mu\widetilde{x}_0, \\
\partial_t \widetilde{e}_l &= -\partial_a \widetilde{e}_l - \mu\widetilde{e}_l - \frac{\sigma}{\epsilon}\widetilde{e}_l - \Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)(\widetilde{e}_w+\widetilde{e}_v) - \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)(\widetilde{e}_w+\widetilde{e}_v) \\
&\quad + \Lambda(\widetilde{e}_l,\widetilde{e}_x,\widetilde{e}_z)(n-\overline{w}-\overline{v}) - \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)(n-\overline{w}-\overline{v}) - \partial_a\widetilde{l}_0 - \mu\widetilde{l}_0, \\
\partial_t \widetilde{e}_v &= -\partial_a \widetilde{e}_v - \mu\widetilde{e}_v - (p+\psi)\widetilde{e}_v - p\widetilde{e}_w,
\end{aligned}
\tag{29}
$$

with boundary condition

$$
\begin{aligned}
&\widetilde{e}_w(0,t)=0, \quad \widetilde{e}_z(0,t)=-\widetilde{z}_0(0,t/\epsilon), \quad \widetilde{e}_x(0,t)=-\widetilde{x}_0(0,t/\epsilon), \\
&\widetilde{e}_l(0,t)=-\widetilde{l}_0(0,t/\epsilon), \quad \widetilde{e}_v(0,t)=0,
\end{aligned}
\tag{30}
$$

as we assumed that the initial condition for the original problem satisfies
$(\overset{\circ}{s},\overset{\circ}{l},\overset{\circ}{i},\overset{\circ}{c},\overset{\circ}{r},\overset{\circ}{v}) \in D(\mathcal{A})$, and initial condition

$$
\widetilde{e}_w(a,0)=\widetilde{e}_z(a,0)=\widetilde{e}_x(a,0)=\widetilde{e}_l(a,0)=\widetilde{e}_v(a,0)=0. \tag{31}
$$

We see that at the boundary we still have terms that are of order lower than ϵ except for $\widetilde{e}_w(0,t)=0$ and $\widetilde{e}_v(0,t)=0$. Fortunately, to eliminate this initial layer contribution on the boundary, we need to introduce the corner layer by simultaneously rescaling time and age according to $\tau=t/\epsilon$ and $\alpha=a/\epsilon$. Unfortunately, the standard approach to the corner layer, [3], will not suffice here as the corner layer equations will not incorporate the multiplication by μ. Thus the classical corner layer will not belong to $D(\mathcal{M})$ and we will not be able to substitute the error terms into the equations, as in (29)–(31). To remedy the problem, we define the corner corrector to be the solution to

$$
\begin{aligned}
\partial_t \breve{z} &= -\partial_a \breve{z} - \mu\breve{z} - \frac{\gamma_1}{\epsilon}\breve{z}, \\
\partial_t \breve{x} &= -\partial_a \breve{x} - \mu\breve{x} - \frac{\gamma_2}{\epsilon}\breve{x} - \frac{\gamma_1}{\epsilon}\cdot\frac{q\sigma}{\sigma-\gamma_1}\breve{l}, \\
\partial_t \breve{l} &= -\partial_a \breve{l} - \mu\breve{l} - \frac{\sigma}{\epsilon}\breve{l},
\end{aligned}
\tag{32}
$$

with boundary condition

$$
\breve{z}(0,t)=-\widetilde{z}_0(0,t/\epsilon), \quad \breve{x}(0,t)=-\widetilde{x}_0(0,t/\epsilon), \quad \breve{l}(0,t)=-\widetilde{l}_0(0,t/\epsilon), \tag{33}
$$

and initial condition

$$
\breve{z}(a,0)=c_\epsilon e^{-\frac{a}{\epsilon}}, \quad \breve{x}(a,0)=d_\epsilon e^{-\frac{a}{\epsilon}}, \quad \breve{l}(a,0)=h_\epsilon e^{-\frac{a}{\epsilon}}, \tag{34}
$$

where c_ϵ, d_ϵ, and h_ϵ are constants obtained from the equality of the boundary and the initial condition at $(a,t)=(0,0)$, such that the classical solvability of the problem with inhomogeneous boundary condition holds. Hence,

$$\begin{aligned} c_\epsilon &= -\widetilde{z}_0(0,0) = -\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(0) - \overset{\circ}{i}(0), \quad h_\epsilon = -\overset{\circ}{l}(0), \\ d_\epsilon &= -\widetilde{x}_0(0,0) = -\frac{q(0)\gamma_1}{\gamma_1-\gamma_2(0)}\cdot\frac{\sigma}{\sigma-\gamma_1}\overset{\circ}{l}(0) - \frac{q(0)\gamma_1}{\gamma_1-\gamma_2(0)}\overset{\circ}{i}(0) - \overset{\circ}{c}(0) \end{aligned} \tag{35}$$

as we assumed that the initial condition for the original problem satisfies
$(\overset{\circ}{s},\overset{\circ}{l},\overset{\circ}{i},\overset{\circ}{c},\overset{\circ}{r},\overset{\circ}{v}) \in D(\mathcal{A})$.

The necessary estimates of the corner layer solution will be provided later. Here, assuming that it is sufficiently regular, we consider the new approximation

$$(w_\epsilon, z_\epsilon, x_\epsilon, l_\epsilon, v_\epsilon) \approx (\overline{w}, \widetilde{z}_0 + \breve{l}, \widetilde{x}_0 + \breve{i}, \widetilde{l}_0 + \breve{c}, \overline{v}).$$

The corresponding error,

$$\begin{aligned} \breve{E} &= (\breve{e}_w, \breve{e}_z, \breve{e}_x, \breve{e}_l, \breve{e}_v) \\ &= (w_\epsilon - \overline{w}, z_\epsilon - \widetilde{z}_0 - \breve{z}, x_\epsilon - \widetilde{x}_0 - \breve{x}, l_\epsilon - \widetilde{l}_0 - \breve{l}, v_\epsilon - \overline{v}) \\ &= (\widetilde{e}_w, \widetilde{e}_z - \breve{z}, \widetilde{e}_x - \breve{x}, \widetilde{e}_l - \breve{l}, \widetilde{e}_v), \end{aligned}$$

satisfies the system

$$\begin{aligned} \partial_t\breve{e}_w &= -\partial_a\breve{e}_w - \mu\breve{e}_w - \Lambda(\breve{e}_l,\breve{e}_x,\breve{e}_z)\,(\breve{e}_w+\breve{e}_v) - \Lambda(\breve{l},\breve{x},\breve{z})\,(\breve{e}_w+\breve{e}_v) \\ &\quad + \Lambda(\breve{e}_l,\breve{e}_x,\breve{e}_z)\,(n-\overline{w}-\overline{v}) - \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(\breve{e}_w+\breve{e}_v) \\ &\quad - \Lambda(\breve{l},\breve{x},\breve{z})\breve{w} + \Lambda(\breve{l},\breve{x},\breve{z})\,(n-\overline{w}-\overline{v}) \\ &\quad + \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(n-\overline{w}-\overline{v}), \\ \partial_t\breve{e}_z &= -\partial_a\breve{e}_z - \mu\breve{e}_z - \frac{\gamma_1}{\epsilon}\breve{e}_z - \frac{\sigma}{\sigma-\gamma_1}\Lambda(\breve{e}_l,\breve{e}_x,\breve{e}_z)\,(\breve{e}_w+\breve{e}_v) \\ &\quad - \frac{\sigma}{\sigma-\gamma_1}\Lambda(\breve{l},\breve{x},\breve{z})\,(\breve{e}_w+\breve{e}_v) + \frac{\sigma}{\sigma-\gamma_1}\Lambda(\breve{e}_l,\breve{e}_x,\breve{e}_z)\,(n-\overline{w}-\overline{v}) \\ &\quad - \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(\breve{e}_w+\breve{e}_v) + \frac{\sigma}{\sigma-\gamma_1}\Lambda(\breve{l},\breve{x},\breve{z})\,(n-\overline{w}-\overline{v}) \\ &\quad + \frac{\sigma}{\sigma-\gamma_1}\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\,(n-\overline{w}-\overline{v}) - \partial_a\widetilde{z}_0 - \mu\widetilde{z}_0, \end{aligned}$$

$$
\begin{aligned}
\partial_t \breve{e}_x &= -\partial_a \breve{e}_x - \mu \breve{e}_x - \frac{\gamma_2}{\epsilon}\breve{e}_x - \frac{1}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma-\gamma_1}\breve{e}_l + \frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\gamma_2^{'}}{\gamma_1-\gamma_2}\breve{e}_z \\
&\quad - \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}\Big(\Lambda(\breve{e}_l,\breve{e}_x,\breve{e}_z)+\Lambda(\breve{l},\breve{x},\breve{z})\Big)(\breve{e}_w+\breve{e}_v) \\
&\quad + \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}\Big(\Lambda(\breve{e}_l,\breve{e}_x,\breve{e}_z)(n-\overline{w}-\overline{v})-\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)(\breve{e}_w+\breve{e}_v)\Big) \\
&\quad + \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}\Big(\Lambda(\breve{l},\breve{x},\breve{z})+\Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)\Big)(n-\overline{w}-\overline{v}) \\
&\quad + \frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\gamma_2^{'}}{\gamma_1-\gamma_2}(\breve{z}+\widetilde{z}_0) - \frac{1}{\epsilon}\cdot\frac{\sigma\gamma_1}{\sigma-\gamma_1}\breve{l} - \partial_a\widetilde{x}_0 - \mu\widetilde{x}_0, \\
\partial_t \breve{e}_l &= -\partial_a \breve{e}_l - \mu\breve{e}_l - \frac{\sigma}{\epsilon}\breve{e}_l - \Lambda(\breve{e}_l,\breve{e}_x,\breve{e}_z)(\breve{e}_w+\breve{e}_v) - \Lambda(\breve{l},\breve{x},\breve{z})(\breve{e}_w+\breve{e}_v) \\
&\quad + \Big(\Lambda(\breve{e}_l,\breve{e}_x,\breve{e}_z)+\Lambda(\breve{l},\breve{x},\breve{z})\Big)(n-\overline{w}-\overline{v}) - \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)(\breve{e}_w+\breve{e}_v) \\
&\quad + \Lambda(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0)(n-\overline{w}-\overline{v}) - \partial_a\widetilde{l}_0 - \mu\widetilde{l}_0, \\
\partial_t \breve{e}_v &= -\partial_a \breve{e}_v - \mu\breve{e}_v - (p+\psi)\breve{e}_v - p\breve{e}_w,
\end{aligned}
\tag{36}
$$

with boundary condition

$$
\breve{e}_w(0,t) = \breve{e}_z(0,t) = \breve{e}_x(0,t) = \breve{e}_l(0,t) = \breve{e}_v(0,t) = 0, \tag{37}
$$

and initial condition

$$
\begin{aligned}
&\breve{e}_w(a,0) = 0, \quad \breve{e}_z(a,0) = -c_\epsilon e^{-\frac{a}{\epsilon}}, \quad \breve{e}_x(a,0) = -d_\epsilon e^{-\frac{a}{\epsilon}}, \\
&\breve{e}_l(a,0) = -h_\epsilon e^{-\frac{a}{\epsilon}}, \quad \breve{e}_v(a,0) = 0.
\end{aligned}
\tag{38}
$$

4. Corner corrector estimates

Lemma 4.1 *Let $M_\theta = \mu + \theta/\epsilon$, where θ is a constant. If η is the solution of*

$$
\partial_t\eta = -\partial_a\eta - M_\theta\eta, \quad \eta(0,t) = -\delta_1 e^{-\frac{\theta}{\epsilon}t}, \quad \eta(a,0) = -\delta_2 e^{-\frac{a}{\epsilon}},
$$

then there exists a nonzero constant C_η such that

$$
\|\eta(t)\| \le \epsilon C_\eta e^{-\frac{\theta}{2\epsilon}t}. \tag{39}
$$

Proof The solution of (39) is given by

$$
\eta(a,t) = \begin{cases} \delta_2 e^{-\frac{a-t}{\epsilon}}\dfrac{\Pi_{M_\theta}(a)}{\Pi_{M_\theta}(a-t)} & \text{if } a > t, \\ \delta_1 e^{-\frac{\theta}{\epsilon}(t-a)}\Pi_{M_\theta}(a) & \text{if } a < t, \end{cases}
$$

where Π_{M_θ} is defined by (9) with μ replaced by M_θ. Using the above expression of $\eta(a,t)$, we get

$$\begin{aligned}
\|\eta(t)\| &\leq |\delta| \left(\int_0^{\min\{t,a_+\}} e^{-\frac{\theta}{2\epsilon}(t-a)} \Pi_{M_\theta}(a)\, da + e^{\frac{t}{\epsilon}} \int_{\min\{t,a_+\}}^{a_+} e^{-\frac{a}{\epsilon}} \frac{\Pi_{M_\theta}(a)}{\Pi_{M_\theta}(a-t)}\, da \right) \\
&\leq |\delta| e^{-\frac{\theta}{2\epsilon}t} \left(\int_0^{\min\{t,a_+\}} e^{-\frac{\theta}{2\epsilon}a} e^{-\int_0^a \mu(\tau)\, d\tau}\, da + e^{\frac{t}{\epsilon}} \int_{\min\{t,a_+\}}^{a_+} e^{-\frac{a}{\epsilon} - \int_{a-t}^a \mu(\tau)\, d\tau}\, da \right) \\
&\leq |\delta| e^{-\frac{\theta}{2\epsilon}t} \left(\int_0^{\min\{t,a_+\}} e^{-\frac{\theta}{2\epsilon}a}\, da + e^{\frac{t}{\epsilon}} \int_{\min\{t,a_+\}}^{a_+} e^{-\frac{a}{\epsilon}}\, da \right) \\
&\leq \epsilon e^{-\frac{\theta}{2\epsilon}t} |\delta| \left(\frac{2}{\theta} + e^{-\frac{\min\{t,a_+\}-t}{\epsilon}} - e^{-\frac{a_+-t}{\epsilon}} \right) = \epsilon K_\theta e^{-\frac{\theta}{2\epsilon}t},
\end{aligned}$$

where $|\delta| = \max\{|\delta_1|, |\delta_2|\}$, and we arrive at (39) □

Now we want to specify (39) to $\breve{z}$ and $\breve{l}$, solution to (32)–(34), where $\delta_1 = \delta_2$. From Lemma 4.1, we have the following result:

Proposition 4.2 *Let $\breve{z}$ and $\breve{l}$ be the solution to (32)–(34). Then there are constants K_z and K_l, depending on the coefficients and the $\mathrm{W}^{1,1}(\mathbb{R}_+)$ norm of the initial condition $\overset{\circ}{i}$ and $\overset{\circ}{l}$, such that for any $t \in \mathbb{R}_+$,*

$$\|\breve{z}(t)\| \leq \epsilon K_z e^{-\frac{\gamma_1}{2\epsilon}t}, \tag{40}$$

$$\|\breve{l}(t)\| \leq \epsilon K_l e^{-\frac{\sigma}{2\epsilon}t}, . \tag{41}$$

Moreover, we also have the following result:

Proposition 4.3 *Let $0 < \underline{\gamma_2} < \gamma_2 < \overline{\gamma_2} < \gamma_1 < \sigma$. Let $\breve{x}_1$ be the solution to (32)–(34). Then there is a constant K_x, depending on the coefficients and on the $\mathrm{W}^{1,1}([0,a_+])$ norm of the initial conditions $\overset{\circ}{c}$, $\overset{\circ}{i}$ and $\overset{\circ}{l}$, such that for any $t \in \mathbb{R}_+$,*

$$\|\breve{x}(t)\| \leq \epsilon K_x e^{-\frac{\gamma_2}{\epsilon}t}. \tag{42}$$

Proof Let $\breve{x}$ be the solution to (32)–(34). We get

$$\breve{x}(a,t) = \begin{cases} \zeta_1 e^{-\frac{a-t}{\epsilon}} \dfrac{\Pi_{M_{\gamma_2}}(a)}{\Pi_{M_{\gamma_2}}(a-t)} + \dfrac{\Pi_{M_{\gamma_2}}(a)}{\epsilon} \displaystyle\int_0^t \frac{F(a-t+\tau,\tau)}{\Pi_{M_{\gamma_2}}(a-t+\tau)}\, d\tau & \text{if } a > t, \\ \zeta_2 e^{-\frac{\gamma_2(a)}{\epsilon}(t-a)} \Pi_{M_{\gamma_2}}(a) + \dfrac{\Pi_{M_{\gamma_2}}(a)}{\epsilon} \displaystyle\int_0^a \frac{F(\tau,t-a+\tau)}{\Pi_{M_{\gamma_2}}(\tau)}\, d\tau & \text{if } a < t, \end{cases}$$

where $\zeta_1 = -\widetilde{x}_0(0,0)$, $\zeta_2 = -\widetilde{x}_0(0,0) - \dfrac{\sigma}{\sigma-\gamma_1} \cdot \dfrac{\overset{\circ}{l}(0)}{\sigma-\gamma_2(0)} \left(e^{-\frac{\sigma-\gamma_2(0)}{\epsilon}(t-a)} - 1 \right)$ and $F(a,t) = -\dfrac{\sigma\gamma_1}{\sigma-\gamma_1} q(a) \breve{l}(a,t)$.

It follows that

$$\begin{aligned}
\|\breve{x}(t)\| &\leq |\zeta| \left(\int_0^{\min\{t,a_+\}} e^{-\frac{\gamma_2}{2\epsilon}(t-a)} \Pi_{M_{\gamma_2}}(a)\, da + e^{\frac{t}{\epsilon}} \int_{\min\{t,a_+\}}^{a_+} e^{-\frac{a}{\epsilon}} \frac{\Pi_{M_{\gamma_2}}(a)}{\Pi_{M_{\gamma_2}}(a-t)}\, da \right) \\
&\quad + \frac{1}{\epsilon} \int_0^{\min\{t,a_+\}} \Pi_{M_{\gamma_2}}(a) \left(\int_0^a \frac{|F(\tau, t-a+\tau)|}{\Pi_{M_{\gamma_2}}(\tau)}\, d\tau \right) da \\
&\quad + \frac{1}{\epsilon} \int_{\min\{t,a_+\}}^{a_+} \Pi_{M_{\gamma_2}}(a) \left(\int_0^t \frac{|F(a-t+\tau, \tau)|}{\Pi_{M_{\gamma_2}}(a-t+\tau)}\, d\tau \right) da \\
&\leq |\zeta| e^{-\frac{\gamma_2}{2\epsilon}t} \left(\int_0^{\min\{t,a_+\}} e^{-\frac{\gamma_2}{2\epsilon}a} e^{-\int_0^a \mu(\tau)\, d\tau}\, da + e^{\frac{t}{\epsilon}} \int_{\min\{t,a_+\}}^{a_+} e^{-\frac{a}{\epsilon} - \int_{a-t}^a \mu(\tau)\, d\tau}\, da \right) \\
&\quad + \frac{\sigma\overline{q}\gamma_1}{\sigma-\gamma_1} \int_0^{\min\{t,a_+\}} e^{-\frac{\gamma_2}{\epsilon}a} \left(\int_0^a e^{\frac{\gamma_2}{\epsilon}\tau} |\breve{l}(\tau, t-a+\tau)|\, d\tau \right) da \\
&\quad + \frac{\sigma\overline{q}\gamma_1}{\sigma-\gamma_1} \int_{\min\{t,a_+\}}^{a_+} e^{-\frac{\gamma_2}{\epsilon}t} \left(\int_0^t e^{\frac{\gamma_2}{\epsilon}\tau} |\breve{l}(a-t+\tau, \tau)|\, d\tau \right) da \\
&\leq |\zeta| e^{-\frac{\gamma_2}{2\epsilon}t} \left(\int_0^{\min\{t,a_+\}} e^{-\frac{\gamma_2}{2\epsilon}a}\, da + e^{\frac{t}{\epsilon}} \int_{\min\{t,a_+\}}^{a_+} e^{-\frac{a}{\epsilon}}\, da \right) \\
&\quad + \frac{\sigma\overline{q}\gamma_1}{\sigma-\gamma_1} \int_0^{\min\{t,a_+\}} e^{\frac{\gamma_2}{\epsilon}\tau} |\breve{l}(\tau, t-a+\tau)| \left(\int_\tau^{\min\{t,a_+\}} e^{-\frac{\gamma_2}{\epsilon}a}\, da \right) d\tau \\
&\quad + \frac{\sigma\overline{q}\gamma_1}{\sigma-\gamma_1} \int_{a-t}^{a_+} e^{\frac{\gamma_2}{\epsilon}\tau} |\breve{l}(\tau, t-a+\tau)| \left(\int_\tau^{a_+} e^{-\frac{\gamma_2}{\epsilon}a}\, da \right) d\tau \\
&\leq \epsilon e^{-\frac{\gamma_2}{2\epsilon}} |\zeta| \left(\frac{2}{\underline{\gamma_2}} + e^{-\frac{\min\{t,a_+\}-t}{\epsilon}} - e^{-\frac{a_+-t}{\epsilon}} \right) + \frac{2\sigma\overline{q}\gamma_1}{\sigma-\gamma_1} \|\breve{l}(t)\| \\
&= \epsilon K_{\gamma_2} e^{-\frac{\gamma_2}{2\epsilon}} + \frac{2\sigma\overline{q}\gamma_1}{\sigma-\gamma_1} \|\breve{l}(t)\|,
\end{aligned}$$

where $|\zeta| = \max\{|\zeta_1|, |\zeta_2|\}$. Using (41), we get

$$\begin{aligned}
\|\breve{x}(t)\| &\leq \epsilon K_{\gamma_2} e^{-\frac{\gamma_2}{2\epsilon}t} + \epsilon \frac{2\sigma\overline{q}\gamma_1}{\sigma-\gamma_1} K_l e^{-\frac{\sigma}{2\epsilon}t} \\
&\leq \epsilon \left(K_{\gamma_2} + \frac{2\sigma\overline{q}\gamma_1}{\sigma-\gamma_1} K_l \right) e^{-\frac{\gamma_2}{2\epsilon}t} = \epsilon K_x e^{-\frac{\gamma_2}{2\epsilon}t}.
\end{aligned}$$

□

5. Main result

Theorem 5.1 *Let the coefficients of the problem (6) satisfy* **A1**–**A4** *together with (11), and* $(\overset{\circ}{s},\overset{\circ}{l},\overset{\circ}{i},\overset{\circ}{c},\overset{\circ}{r},\overset{\circ}{v})$ *be such that* $(s_\epsilon,l_\epsilon,i_\epsilon,c_\epsilon,r_\epsilon,v_\epsilon)$ *is a classical solution to (6). Then there exist constants* C_1,C_2,C_3,C_4,C_5,C_6, *depending only on the coefficients of the problem and the* $D(\mathcal{S})\cap D(\mathcal{M})$ *norm of the initial conditions, such that for all sufficiently small* $\epsilon>0$

$$\|s_\epsilon(t)-(n(t)-\overline{w}(t)-\overline{v}(t))\|\le \epsilon C_1, \tag{43}$$

$$\|l_\epsilon(t)-\overset{\circ}{l}e^{-\frac{\sigma}{\epsilon}t}\|\le \epsilon C_2, \tag{44}$$

$$\left\|i_\epsilon(t)-\left(\overset{\circ}{i}e^{-\frac{\gamma_1}{\epsilon}t}+\frac{\sigma\overset{\circ}{l}}{\sigma-\gamma_1}\left(e^{-\frac{\gamma_1}{\epsilon}t}-e^{-\frac{\sigma}{\epsilon}t}\right)\right)\right\|\le \epsilon C_3, \tag{45}$$

$$\begin{aligned}\left\|c_\epsilon(t)-\left(\overset{\circ}{c}e^{-\frac{\gamma_2}{\epsilon}t}-\frac{q\gamma_1^2}{(\gamma_1-\gamma_2)^2}\left(\frac{\sigma\overset{\circ}{l}}{\sigma-\gamma_1}+\overset{\circ}{i}\right)\left(e^{-\frac{\gamma_2}{\epsilon}t}-e^{-\frac{\gamma_1}{\epsilon}t}\right)\right.\right.\\ \left.\left.-\frac{\sigma\overset{\circ}{l}}{(\sigma-\gamma_1)(\sigma-\gamma_2)}\left(e^{-\frac{\gamma_2}{\epsilon}t}-e^{-\frac{\sigma}{\epsilon}t}\right)\right)\right\|\le \epsilon C_4,\end{aligned} \tag{46}$$

$$\begin{aligned}\left\|r_\epsilon(t)-\left(\overline{w}-\frac{\gamma_1\overset{\circ}{l}}{\sigma-\gamma_1}e^{-\frac{\sigma}{\epsilon}t}-\frac{(1-q)\gamma_1-\gamma_2}{\gamma_1-\gamma_2}\left(\frac{\sigma\overset{\circ}{l}}{\sigma-\gamma_1}+\overset{\circ}{i}\right)e^{-\frac{\gamma_1}{\epsilon}t}\right.\right.\\ -\frac{\sigma\overset{\circ}{l}}{(\sigma-\gamma_1)(\gamma_1-\gamma_2)}\left(e^{-\frac{\gamma_2}{\epsilon}t}-e^{-\frac{\sigma}{\epsilon}t}\right)\\ -\left(\frac{q\sigma\gamma_1\overset{\circ}{l}}{(\sigma-\gamma_1)(\gamma_1-\gamma_2)}+\frac{q\gamma_1\overset{\circ}{i}}{\gamma_1-\gamma_2}+\overset{\circ}{c}\right)e^{-\frac{\gamma_1}{\epsilon}t}\\ \left.\left.-\frac{q\gamma_1^2}{(\gamma_1-\gamma_2)^2}\left(\frac{\sigma\overset{\circ}{l}}{\sigma-\gamma_1}+\overset{\circ}{i}\right)\left(e^{-\frac{\gamma_2}{\epsilon}t}-e^{-\frac{\gamma_1}{\epsilon}t}\right)\right)\right\|\le \epsilon C_5,\end{aligned} \tag{47}$$

$$\|v_\epsilon(t)-\overline{v}(t)\|\le \epsilon C_6. \tag{48}$$

Proof We shall simplify the notation and subsequent calculations by introducing the rescaled errors $\breve{e}_w=\epsilon d$, $\breve{e}_z=\epsilon f$, $\breve{e}_x=\epsilon g$, $\breve{e}_l=\epsilon h$ and $\breve{e}_v=\epsilon j$; that is,

$$\begin{aligned}&w_\epsilon=\overline{w}+\epsilon d,\quad z_\epsilon=\widetilde{z}_0+\breve{z}+\epsilon f,\quad x_\epsilon=\widetilde{x}_0+\breve{x}+\epsilon g,\\ &l_\epsilon=\widetilde{l}_0+\breve{l}+\epsilon h,\quad v_\epsilon=\overline{v}+\epsilon j.\end{aligned} \tag{49}$$

This converts the system (36)–(38) into

$$\begin{aligned}\partial_t d=&-\partial_a d-\mu d-(d+j)F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z})-\epsilon(d+j)\Lambda(h,g,f)\\ &+\Lambda(h,g,f)G(n,\overline{w},\overline{v})+\frac{1}{\epsilon}F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z})G(n,\overline{w},\overline{v}),\end{aligned}$$

$$
\begin{aligned}
\partial_t f &= -\partial_a f - \mu f - \frac{\gamma_1}{\epsilon} f - \epsilon \frac{\sigma}{\sigma-\gamma_1}(d+j)\Lambda(h,g,f) - \frac{\sigma}{\sigma-\gamma_1}(d+j) \\
&\quad \times F(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0, \breve{l}, \breve{x}, \breve{z}) + \frac{\sigma}{\sigma-\gamma_1}\Lambda(h,g,f)G(n,\overline{w},\overline{v}) \\
&\quad + \frac{1}{\epsilon}\left(-\partial_a \widetilde{z}_0 - \mu \widetilde{z}_0 + \frac{\sigma}{\sigma-\gamma_1}F(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0, \breve{l}, \breve{x}, \breve{z})G(n,\overline{w},\overline{v})\right), \\
\partial_t g &= -\partial_a g - \mu g - \frac{\gamma_2}{\epsilon} g - \frac{1}{\epsilon}\cdot\frac{\sigma}{\sigma-\gamma_1} h + \frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\gamma_2^{'}}{\gamma_1-\gamma_2} f, \\
&\quad - \epsilon\frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}(d+j)\Lambda(h,g,f) - \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2} \\
&\quad \times \left((d+j)F(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0, \breve{l}, \breve{x}, \breve{z}) - \Lambda(h,g,f)G(n,\overline{w},\overline{v})\right) \\
&\quad + \frac{1}{\epsilon}\left(\frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}F(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0, \breve{l}, \breve{x}, \breve{z})G(n,\overline{w},\overline{v})\right. \\
&\quad \left. + \frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\gamma_2^{'}}{\gamma_1-\gamma_2}(\widetilde{z}_0+\breve{z}) - \partial_a\widetilde{x}_0 - \mu\widetilde{x}_0\right) - \frac{1}{\epsilon^2}\cdot\frac{\sigma\gamma_1}{\sigma-\gamma_1}\breve{l}, \\
\partial_t h &= -\partial_a h - \mu h - \frac{\sigma}{\epsilon} h - \epsilon(d+j)\Lambda(h,g,f) - (d+j)F(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0, \breve{l}, \breve{x}, \breve{z}) \\
&\quad + \Lambda(h,g,f)G(n,\overline{w},\overline{v}) + \frac{1}{\epsilon}\Big(- \partial_a\widetilde{l}_0 - \mu\widetilde{l}_0 \\
&\quad F(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0, \breve{l}, \breve{x}, \breve{z})G(n,\overline{w},\overline{v})\Big), \\
\partial_t j &= -\partial_a j - \mu j - (p+\psi)j - pd,
\end{aligned}
\tag{50}
$$

with boundary condition

$$
d(0,t) = f(0,t) = g(0,t) = h(0,t) = j(0,t) = 0, \tag{51}
$$

and with initial condition

$$
\begin{aligned}
&d(a,0) = 0, \quad f(a,0) = -\frac{c_\epsilon}{\epsilon}e^{-\frac{a}{\epsilon}}, \quad g(a,0) = -\frac{d_\epsilon}{\epsilon}e^{-\frac{a}{\epsilon}}, \\
&h(a,0) = -\frac{h_\epsilon}{\epsilon}e^{-\frac{a}{\epsilon}}, \quad j(a,0) = 0,
\end{aligned}
\tag{52}
$$

where

$$
\begin{aligned}
F(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0, \breve{l}, \breve{x}, \breve{z}) &= \Lambda(\widetilde{l}_0, \widetilde{x}_0, \widetilde{z}_0) + \Lambda(\breve{l}, \breve{x}, \breve{z}), \\
G(n,\overline{w},\overline{v}) &= n - \overline{w} - \overline{v}.
\end{aligned}
$$

In the sequel we consider (11), as mentioned earlier. In this case $\lambda_\mu \leq 0$ and n and $\overline{w}$ satisfy (10). Furthermore, we shall need the estimate for the $\overline{v}$ solution to the scalar McKendrick problem (13). For this,

we first consider the auxiliary problem

$$\begin{aligned} \partial_t \phi &= -\partial_a \phi - \mu\phi - \psi\phi, \\ \phi(0,t) &= (1-\omega)\int_0^{a_+} \beta(a)\phi(a)\,da, \\ \phi(a,0) &= \overset{o}{\phi}(a). \end{aligned} \tag{53}$$

Using the fact that $\left(e^{tA}\right)_{t\geq 0}$ is a positive semigroup, we get

$$\|\phi(t)\| \leq Me^{-\psi t}\|\overset{o}{\phi}\|, \quad t \geq 0. \tag{54}$$

Next we apply the Duhamel formula to the Equation (13). Taking the norm and using (54) lead to

$$\begin{aligned} e^{\psi t}\|\overline{v}(t)\| \leq M\|\overset{o}{v}\| &+ \int_0^{a_+}\left(\int_0^t e^{\psi s}|p(a,s)||\overline{v}(a,s)|\,ds\right)da \\ &+ \int_0^{a_+}\left(\int_0^t e^{\psi s}|p(a,s)|\Big(|n(a,s)|+|\overline{w}(a,s)|\Big)\,ds\right)da. \end{aligned} \tag{55}$$

Since $p \in \mathcal{C}\left([0,a_+]\times[0,T]\right)$, see **A4**, we can find a positive real number δ_0 such that if $0 < a+t < \delta_0$ then $|p(a,t)-p(0,0)| < \epsilon_0$. We see that

$$|p(a,t)| \leq \epsilon_0 + c_0, \tag{56}$$

where $c_0 = |p(0,0)|$ and hence, from (55), we get

$$\begin{aligned} e^{\psi t}\|\overline{v}(t)\| \leq M\|\overset{o}{v}\| &+ (\epsilon_0+c_0)\int_0^t e^{\psi s}\|\overline{v}(s)\|\,ds \\ &+ (\epsilon_0+c_0)\int_0^t e^{\psi s}\Big(\|n(s)\|+\|\overline{w}(s)\|\Big)\,ds. \end{aligned}$$

This leads, by (10), to

$$\begin{aligned}
e^{\psi t}\|\overline{v}(t)\| &\le M\|\overset{\circ}{v}\| + (\epsilon_0 + c_0)\int_0^t e^{\psi s}\|\overline{v}(s)\|\,ds \\
&\quad + M(\epsilon_0 + c_0)\int_0^t e^{\psi s}\Big(\|\overset{\circ}{n}\| + \|\overset{\circ}{w}\|\Big)\,ds, \\
&\le M\left(1 + (\epsilon_0 + c_0)\int_0^t e^{\psi s}\,ds\right)\|\overset{\circ}{n}\| \\
&\quad + (\epsilon_0 + c_0)\int_0^t e^{\psi s}\|\overline{v}(s)\|\,ds, \\
&\le M(1 + \epsilon_0 + c_0)e^{\psi t}\|\overset{\circ}{n}\| + (\epsilon_0 + c_0)\int_0^t e^{\psi s}\|\overline{v}(s)\|\,ds
\end{aligned}$$

and the Gronwall inequality gives

$$\|\overline{v}(t)\| \le M(1 + \epsilon_0 + c_0)\left(1 + \frac{\epsilon_0 + c_0}{\epsilon_0 + c_0 - \psi}e^{(\epsilon_0 + c_0 - \psi)t}\right)\|\overset{\circ}{n}\| \le K_v\|\overset{\circ}{n}\|, \tag{57}$$

with $\psi \neq \epsilon_0 + c_0$.

Then, (49), by (10) with (17), (27), (40), (41), (42), and (57), yields

$$\|\epsilon d(t)\| \le C_d\|\overset{\circ}{n}\|_{\mathrm{W}^{1,1}([0,a_+])}, \tag{58}$$

$$\|\epsilon f(t)\| \le C_f\|\overset{\circ}{n}\|_{\mathrm{W}^{1,1}([0,a_+])}\left(1 + e^{-\frac{\gamma_1}{\epsilon}t} + \epsilon e^{-\frac{\gamma_1}{2\epsilon}t}\right), \tag{59}$$

$$\|\epsilon g(t)\| \le C_g\|\overset{\circ}{n}\|_{\mathrm{W}^{1,1}([0,a_+])}\left(1 + e^{-\frac{\gamma_2}{\epsilon}t} + \epsilon e^{-\frac{\gamma_2}{2\epsilon}t}\right), \tag{60}$$

$$\|\epsilon h(t)\| \le C_h\|\overset{\circ}{n}\|_{\mathrm{W}^{1,1}([0,a_+])}\left(1 + e^{-\frac{\sigma}{\epsilon}t} + \epsilon e^{-\frac{\sigma}{2\epsilon}t}\right), \tag{61}$$

$$\|\epsilon j(t)\| \le C_j\|\overset{\circ}{n}\|_{\mathrm{W}^{1,1}([0,a_+])}, \tag{62}$$

for some constants C_d, C_f, C_g, C_h, and C_j.

For any function of a, or a constant, θ, we consider the following auxiliary problem:

$$\partial_t\varrho = -\partial_a\varrho - \mu\varrho - \frac{\theta}{\epsilon}\varrho, \quad \varrho(0,t) = 0, \quad \varrho(a,0) = \overset{\circ}{\varrho}(a). \tag{63}$$

Using the fact that $\left(e^{tA}\right)_{t\ge 0}$ is a positive semigroup, it is easy to see that then

$$\|\varrho(t)\| \le Me^{-\frac{\theta}{\epsilon}t}\|\overset{\circ}{\varrho}\|, \qquad t \ge 0, \tag{64}$$

where, as mentioned earlier, $\underline{\theta} = \min_{a\in[0,a_+]} \theta(a)$. Note that the assumption on R_μ is also sufficient here since $\omega \in [0,1]$. Using this estimate and the Duhamel formula, from (50), we have

$$\begin{aligned}
\|d(t)\| \leq M \int_0^t \Big\| & -(d+j)F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z}) - \epsilon(d+j)\Lambda(h,g,f) \\
& +\Lambda(h,g,f)G(n,\overline{w},\overline{v}) + \frac{1}{\epsilon}F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z})G(n,\overline{w},\overline{v})\Big\| \, ds, \qquad (65)
\end{aligned}$$

$$\begin{aligned}
\|f(t)\| \leq & Me^{-\frac{\gamma_1}{\epsilon}t}|c_\epsilon| + M\int_0^t e^{-\frac{\gamma_1}{\epsilon}(t-s)} \Big\| -\epsilon\frac{\sigma}{\sigma-\gamma_1}(d+j)\Lambda(h,g,f) \\
& - \frac{\sigma}{\sigma-\gamma_1}(d+j)F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z}) + \frac{\sigma}{\sigma-\gamma_1}\Lambda(h,g,f)G(n,\overline{w},\overline{v}) \\
& + \frac{1}{\epsilon}\left(-\partial_a\widetilde{z}_0 - \mu\widetilde{z}_0 + \frac{\sigma}{\sigma-\gamma_1}F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z})G(n,\overline{w},\overline{v})\right)\Big\| \, ds, \qquad (66)
\end{aligned}$$

$$\begin{aligned}
\|g(t)\| \leq & Me^{-\frac{\gamma_2}{\epsilon}t}|d_\epsilon| + M\int_0^t e^{-\frac{\gamma_2}{\epsilon}(t-s)} \Big\| -\frac{1}{\epsilon}\cdot\frac{\sigma}{\sigma-\gamma_1}h + \frac{q\gamma_1}{\gamma_1-\gamma_2}\cdot\frac{\gamma_2'}{\gamma_1-\gamma_2}f \\
& - \epsilon\frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}(d+j)\Lambda(h,g,f) - \frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2} \\
& \times\left((d+j)F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z}) - \Lambda(h,g,f)G(n,\overline{w},\overline{v})\right) \\
& + \frac{1}{\epsilon}\left(\frac{\sigma}{\sigma-\gamma_1}\cdot\frac{q\gamma_1}{\gamma_1-\gamma_2}F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z})G(n,\overline{w},\overline{v}) + \frac{q\gamma_1}{\gamma_1-\gamma_2}\right. \\
& \left.\times\frac{\gamma_2'}{\gamma_1-\gamma_2}(\widetilde{z}_0+\breve{z}) - \partial_a\widetilde{x}_0 - \mu\widetilde{x}_0\right) - \frac{1}{\epsilon^2}\cdot\frac{\sigma\gamma_1}{\sigma-\gamma_1}\breve{l}\Big\| \, ds, \qquad (67)
\end{aligned}$$

$$\begin{aligned}
\|h(t)\| \leq & Me^{-\frac{\sigma}{\epsilon}t}|h_\epsilon| + M\int_0^t e^{-\frac{\sigma}{\epsilon}(t-s)} \Big\| - \epsilon(d+j)\Lambda(h,g,f) \\
& - (d+j)F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z}) + \Lambda(h,g,f)G(n,\overline{w},\overline{v}) \\
& + \frac{1}{\epsilon}\left(-\partial_a\widetilde{l}_0 - \mu\widetilde{l}_0 + F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z})G(n,\overline{w},\overline{v})\right)\Big\| \, ds, \qquad (68)
\end{aligned}$$

$$\|j(t)\| \leq M\int_0^t e^{-\psi(t-s)}\,\|-pj-pd\|\,ds. \qquad (69)$$

Next, by (59), (60), and (61), we obtain

$$\|\epsilon\Lambda(h,g,f)\| \leq k_1\left(1 + e^{-\frac{\gamma_1}{\epsilon}t} + e^{-\frac{\gamma_2}{\epsilon}t} + e^{-\frac{\sigma}{\epsilon}t} + \epsilon e^{-\frac{\gamma_1}{\epsilon}t} + \epsilon e^{-\frac{\gamma_2}{\epsilon}t} + \epsilon e^{-\frac{\sigma}{\epsilon}t}\right). \qquad (70)$$

Further, by (27), (40), (41), and (42),

$$\|F(\widetilde{l}_0,\widetilde{x}_0,\widetilde{z}_0,\breve{l},\breve{x},\breve{z})\| \leq k_2 e^{-\frac{\gamma_2}{\epsilon}t}(1+\epsilon)\left(1+e^{-\frac{\gamma_1-\gamma_2}{\epsilon}t}+e^{-\frac{\sigma-\gamma_2}{\epsilon}t}\right). \tag{71}$$

Next, by (10) and (57),

$$\|G(n,\overline{w},\overline{v})\| \leq k_3. \tag{72}$$

Finally, by (27),

$$\|\partial_a\widetilde{l}_0+\mu\widetilde{l}_0\| \leq k_4 e^{-\frac{\sigma}{\epsilon}t}, \quad \|\partial_a\widetilde{x}_0+\mu\widetilde{x}_0\| \leq k_5 e^{-\frac{\gamma_2}{\epsilon}t}, \quad \|\partial_a\widetilde{z}_0+\mu\widetilde{z}_0\| \leq k_6 e^{-\frac{\gamma_1}{\epsilon}t}, \tag{73}$$

where k_4, k_5, and k_6 depend, in particular, on L^1 norm of $\partial_a\overset{\circ}{l}$ and $\mu\overset{\circ}{l}$, $\partial_a\overset{\circ}{x}$ and $\mu\overset{\circ}{x}$ and $\partial_a\overset{\circ}{z}$ and $\mu\overset{\circ}{z}$, respectively; that is, on the $D(\mathcal{A})$ norm of the initial condition.

The error estimates (50)–(52) will be completed below. Firstly, we see that, by (58), (59), (60), (71), (72), and (73) and by defining $H(t)=e^{\frac{\sigma}{\epsilon}t}\|h(t)\|$, (68) can be written as

$$H(t) \leq K_1\int_0^t H(s)\,ds+\Phi_1(t),$$

where K_1 is a constant and $0<\Phi_1(t)\leq k_7(e^{\frac{\sigma}{\epsilon}t}+1)$, and the Gronwall inequality gives

$$H(t) \leq k_7(e^{\frac{\sigma}{\epsilon}t}+1)+K_1k_7e^{K_1t}\int_0^t e^{-K_1s}(e^{\frac{\sigma}{\epsilon}s}+1)\,ds \leq k_8e^{\frac{\sigma}{\epsilon}t}+k_9e^{K_1t}$$

and hence

$$\|h(t)\| \leq k_8+k_9e^{\left(K_1-\frac{\sigma}{\epsilon}\right)t} \leq k_{10}. \tag{74}$$

Secondly, we see that, by (58), (62), (70), (71), (72), (73), and (74) and by defining $F(t)=e^{\frac{\gamma_1}{\epsilon}t}\|f(t)\|$, (66) can be written as

$$F(t) \leq K_2\int_0^t F(s)\,ds+\Phi_2(t),$$

where K_2 is a constant and $0<\Phi_2(t)\leq k_{11}(e^{\frac{\gamma_1}{\epsilon}t}+1)$, and the Gronwall inequality gives

$$F(t) \leq k_{11}(e^{\frac{\gamma_1}{\epsilon}t}+1)+K_2k_{11}e^{K_2t}\int_0^t e^{-K_2s}(e^{\frac{\gamma_1}{\epsilon}s}+1)\,ds \leq k_{12}e^{\frac{\gamma_1}{\epsilon}t}+k_{13}e^{K_1t}$$

and hence

$$\|f(t)\| \leq k_{12}+k_{13}e^{\left(K_1-\frac{\gamma_1}{\epsilon}\right)t} \leq k_{14}. \tag{75}$$

Thirdly, we see that, by (27), (40), (41), (42), (58), (59), (62), (70), (71), (72), (73), and (74) and by defining $G(t)=e^{\frac{\gamma_2}{\epsilon}t}\|g(t)\|$, (67) can be written as

$$G(t) \leq K_3\int_0^t G(s)\,ds+\Phi_3(t),$$

where K_3 is a constant and $0 < \Phi_3(t) \leq k_{15}(e^{\frac{\gamma_2}{\epsilon}t} + 1)$, and the Gronwall inequality gives

$$G(t) \leq k_{15}(e^{\frac{\gamma_2}{\epsilon}t} + 1) + K_3 k_{15} e^{K_3 t} \int_0^t e^{-K_3 s}(e^{\frac{\gamma_2}{\epsilon}s} + 1)\, ds \leq k_{16} e^{\frac{\gamma_2}{\epsilon}t} + k_{17} e^{K_3 t}$$

and hence

$$\|g(t)\| \leq k_{16} + k_{17} e^{(K_3 - \frac{\gamma_2}{\epsilon})t} \leq k_{18}. \tag{76}$$

Next, we see that, by (27), (40), (41), (42), (70), (71), (72), (73), (74), and (75), (65) can be written as

$$\|d(t)\| \leq k_{19} \int_0^t e^{-\frac{\gamma_2}{2\epsilon}s} \|d(s)\|\, ds + \frac{k_{20}}{\epsilon} \int_0^t e^{-\frac{\gamma_2}{2\epsilon}s}\, ds + k_{21}.$$

Taking the maximum of t over $[0, \infty)$, we get

$$\begin{aligned} \max_{t\in[0,\infty)} \|d(t)\| &\leq k_{19} \int_0^\infty e^{-\frac{\gamma_2}{2\epsilon}s} \|d(s)\|\, ds + \frac{k_{20}}{\epsilon} \int_0^\infty e^{-\frac{\gamma_2}{2\epsilon}s}\, ds + k_{21} \\ &\leq \epsilon k_{22} \max_{t\in[0,\infty)} \|d(t)\| + \epsilon k_{23} + k_{21}, \end{aligned}$$

which implies that $\max_{t\in[0,\infty)} \|d(t)\| \leq \dfrac{k_{21} + \epsilon k_{23}}{1 - \epsilon k_{22}} =: k_{24} < \infty$. Hence

$$\|d(t)\| \leq k_{24}. \tag{77}$$

Finally, we see that, by (56) and (77) and by defining $J(t) = e^{\psi t}\|j(t)\|$, (69) can be written as

$$J(t) \leq K_4 \int_0^t J(s)\, ds + \Phi_4(t),$$

where K_4 is a constant and $0 < \Phi_4(t) \leq k_{25}\left(e^{\psi t} - 1\right)$, and the Gronwall inequality gives

$$J(t) \leq k_{25}\left(e^{\psi t} - 1\right) + K_4 k_{25} e^{K_4 t} \int_0^t e^{-K_4 s}\left(e^{\psi s} - 1\right)\, ds$$

and hence

$$\|j(t)\| \leq \frac{\psi k_{25}}{\psi - K_4} =: k_{26}, \tag{78}$$

where $\psi > K_4$. This completes the proof. □

6. Conclusion

In this paper we proposed an age-structured epidemiological model for the transmission of HBV; then an asymptotic analysis of a singularly perturbed for such a model was performed. The Chapman–Enskog procedure, as an asymptotic method, was used for the mathematical analysis. This asymptotic method was developed in [1, 3, 6, 12] for age-structured epidemiological models and in [13] for the Carleman model of the Boltzmann equation, has enabled a systematic aggregation of variables, and has yielded a good approximated formula with layers correctors, namely initial layer and corner layer correctors. Note that a special corner layer equation was used instead of the standard one, and since the corner layer corrector decays exponentially fast in both a/ϵ and t/ϵ, while the initial layer corrector only does so in t/ϵ, this was neglected because of the $\mathrm{L}([0,a_+],\mathbb{R}^6)$-norm and therefore we obtained the approximation in terms of the initial layer corrector.

In summary, the result obtained shows that the solution of the nonlinear problem (6) can be approximated by the solution of five scalar linear problems and explicitly given initial layer corrector, uniformly for any time. The result is very interesting as the error estimates are uniform on the whole infinite time interval, in contrast to the typical result based on the Tikhonov theorem and classical asymptotic expansions.

References

[1] Abdulle A, Banasiak J, Damlamian A, Sango M. Multiple Scales Problems in Biomathematics, Mechanics, Physics and Numerics. Tokyo, Japan: GAKUTO International Series, 2007.

[2] Banasiak J. Mathematical Modelling in One Dimension. Cambridge, UK: Cambridge University Press, 2013.

[3] Banasiak J, Goswami A, Shindin S. Aggregation age and space structured population models: an asymptotic analysis approach. J Evol Eq 2011; 11: 121-154.

[4] Banasiak J, Lachowicz M. Methods of Small Parameter in Mathematical Biology and Others Applications. Switzerland: Mod Simul Sci Eng Tech, 2014.

[5] Banasiak J, Lamb W. Coagulation, fragmentation and growth processes in a size structured population. Disc Cont Dyn Syst 2012; 17: 445–472.

[6] Banasiak J, M'pika Massoukou RY. A singularly perturbed age structured SIRS model with fast recovery. Disc Cont Dyn Sys 2014; 19: 2383-2399.

[7] Cha Y, Iannelli M, Milner F. Existence and uniqueness of endemic states for the age-structured SIR epidemic model. Math Biosci 1998; 150: 117-133.

[8] Iannelli M. Mathematical Theory of Age-Structured Population Dynamics. Pisa, Italy: Consiglio Nazionale delle Ricerche, 1995.

[9] Iannelli M, Milner FA, Pugliese A. Analytical and numerical results for the age-structured S-I-S epidemic model with mixed inter-intracohort transmission. SIAM J Math Anal 1992; 23: 662-688.

[10] Inaba H. A semigroup approach to the strong ergodic theorem of the multi state stable population process. Math Popul Stud 1988; 1: 49-77.

[11] Li XZ, Liu JX. Stability of an age-structured epidemiological model for hepatitis C. J Appl Math Comput 2008; 27: 159-173.

[12] M'pika Massoukou RY, Age-structured models of mathematical epidemiology. PhD, University of KwaZulu-Natal, DBN, South Africa, 2013.

[13] Palczewski A. Exact and Chapman-Enskog solutions for the Carleman model. Math Meth Appl Sc 1984; 6: 417-432.

[14] Prüss J. On the qualitative behaviour of populations with age-specific interactions. J Comp Math Appl 1983; 9: 327-339.

[15] Pugliese A, Tonetto L. Well-posedness of an infinite system of partial differential equations modelling parasitic infection in an age-structured host. Math Anal Appl 2003; 284: 144-164.

[16] Thieme HR. Mathematics in Population Biology. Princeton, NJ, USA: Princeton University Press, 2011.

[17] Webb GF. Theory of Non-Linear Age Dependent Population Dynamics. New York, NY, USA: Marcel Dekker, 1985.

[18] Yannick KT, Ducrot A, Elvis HDD. A model for hepatitis B with chronological and infection Ages. Appl Math Sci 2013; 7: 5977-5993.

[19] Zou L, Ruan S, Zhang W. Modeling the transmission dynamics and control of hepatitis B virus in China. J Theor Biol 2010; 262: 330-338.

[20] Zou L, Ruan S, Zhang W. An age-structured model for transmission dynamics of hepatitis B. SIAM J Appl Math 2010; 70: 3121-3139.

ANNEXE 3

Article :

J. BANASIAK, A. TRAORE, S. SHINDIN, R.Y.MASSOUKOU, *Improving Heun's method for solving non linear age-structured population equations with infinite life span*, Pionner journal of advances in applied mathematics, volume 20, Issue 1, pp. 55--72 (2017)

Pioneer Journal of Advances in Applied Mathematics
Volume 21, Numbers 1-2, 2017, Pages 55-72
This paper is available online at http://www.pspchv.com/content_PJAAM.html

IMPROVING HEUN'S METHOD FOR SOLVING NON LINEAR AGE-STRUCTURED POPULATION EQUATIONS WITH INFINITE LIFE SPAN

J. BANASIAK[1], A. TRAORE[2], S. SHINDIN[1] and R. Y. MASSOUKOU[1]

[1]School of Mathematical Sciences
Kwazulu-Natal University
Cote D'ivoire
e-mail: banasiak@ukzn.ac.za
shindin@ukzn.ac.za
rodrigue@aims.ac.za

[2]Ecole Normale Superieure d'Abidjan
Cote D'ivoire
e-mail: traboubakari@yahoo.fr

Abstract

We develop a modified Heun's method for solving nonlinear age structured population models with finite and infinite life span. Convergence of the scheme is proven, numerical experiments are given in order to demonstrate efficiency of the proposed algorithm.

1. Introduction

In the paper, we deal with age-structured population models of the form:

$$u_t + u_a = -f(a, P(t))u, \quad 0 < t \le T, \quad 0 < a < +\infty, \tag{1a}$$

Received December 4, 2013

2010 Mathematics Subject Classification: 65M70; 65N22.

Keywords and phrases: age structured population models, Heun's method, characteristic lines.

$$u(t, 0) = g\left(t, \int_0^{+\infty} \beta(a, P(t))\, u\, da\right), \tag{1b}$$

$$u(0, a) = u^0(a), \tag{1c}$$

where

$$P(t) = \int_0^{+\infty} u(t, a)\, da. \tag{2}$$

The independent variables t and a represent age and time, respectively. The value $u(t, a)$ is the density of the population at time t with age a, $f(a, P(t))$ is the age-specific mortality modulus, the age-specific fertility modulus is given by $\beta(a, P(t))$. We observe that functions f and β depend on the total population $P(t)$ at time t as defined in (2). The initial and the boundary functions are given by u^0 and g, respectively. An extensive theoretical study of linear and nonlinear age-structured population models, in particular of system (1), can be found in the works of Iannelli [4], [5] and Webb [10] and allowed us to try to know what could happen with the numerical behaviour of the problem. A number of computational methods for system (1) have been developed and analyzed during the past twenty-five years. For an excellent review of these methods see [1] and [3] where an analysis of numerical schemes (consistency, stability, existence and convergence) is carried out by means of a general framework introduced by Lòpez-Marcos and Sanz-Serna [6]. We mention that the three problems, the one we want to threat, [1] and [3] are completely different. The most studied is the one fixed in [1] in which they integrated a problem where the initial condition is compact support and the integration time is finite, in this case the difficulties of the problem are given by the nonlinearities. The second one presented in [3] shows a problem with finite support but the fact of a finite maximum age makes the mortality to be unbounded. The one treated here deals with bounded and unbounded age intervals.

To avoid high expensive computational effort and regularity of hypotheses, we propose a modified Heun method of second order. We mention that our approach is a particular case to the one proposed by Abia and Lòpez-Marcos [2], where general Runge-Kutta methods were used. However, the algorithm of the formers is limited to a finite maximum age and required high regularity order in hypotheses.

The paper is organized as follows: in Section 2, we present our numerical scheme and discuss regularity conditions needed for its analysis. Consistency, stability and convergence are studied in Section 3. Section 4 is devoted to numerical simulations. Some conclusions are made in Section 5.

2. Numerical Scheme

The method. Along each characteristic line $a_c(t) = t + c$ (c is the age at time $t = 0$) equation (1a) takes the form

$$\frac{d}{dt}u(t, a_c(t)) = -f(a_c(t), P(t))u(t, a_c(t)), \quad u(0, c) = u^0(c). \tag{3}$$

The main idea of the method is to replace (1) with finite system of ODEs (3) and then apply a Runge-Kutta method. The technical details are explained below.

First, we introduce computational grids. For a given finite time interval $[0, T]$, $T > 0$, and a total number of time discretization steps N, we set $h = \frac{T}{N}$. Hypothesis (H_1) imply that there exists positive integer L, such that u^0 is compactly supported in $[0, Lh]$. Using h, N and L we define: the grid on the time axis

$$\mathcal{T} = \{t^n : t^n = nh, 0 \le n \le N\},$$

the grid on the age axis at time t^n

$$\mathcal{A}^n = \{a_j : a_j = jh, 0 \le j \le L + n\},$$

and grids on the characteristic lines

$$\mathcal{C} = \{(t^n, a_j) : 0 \le n \le N, 0 \le j \le L + n\}.$$

We note that points (t^n, a_j) and $(t^n + mh, a_j + mh)$, $m \ge 0$, belong to the same characteristic line.

Second, we assume that subscript j refers to the age a_j and superscript n to the time level t^n. Using this notation we define vector

$$U^n = (U_0^n, ..., U_{L+n}^n)^T,$$

whose components U_j^n represent numerical approximation of $u(t^n, a_j)$. Note that dimension of U^n depends on n.

Third, assume that U^n is given, we evaluate U^{n+1} in two sub-steps. In the first sub-step, we compute the components U_j^{n+1}, $1 \le j \le L + n + 1$, by integrating (3) in the interval $[t^n, t^{n+1}]$. In the second sub-step, we determine boundary value U_0^{n+1} using (1b).

In the first sub-step one can apply arbitrary Runge-Kutta method. However, to advance the numerical solution by a single time step one has to compute several off-grid stage values. This might be expensive as each stage value requires approximation of the integrals $P(t)$. The cost of computations is significantly reduced if stage values coincide with the computational grid. In the paper we employ Heun's method. It is the simplest explicit second order Runge-Kutta method which satisfies the condition mentioned above.

In both sub-steps we have to approximate $P(t^n)$ and $P(t^{n+1})$. The simplest way to do this is to apply a quadrature formula with nodes in $\mathcal{A}^n$. In our case we use composite Simpson's rule (see [9] for a discussion of quadratures in the context of numerical integration of (1)).

One step of the algorithm reads:

$$Q^n = \sum_{j=0}^{L+n} b_j^n U_j^n, \tag{4a}$$

$$F_{1,j}^n = -f(a_j, Q^n) U_j^n, \quad 0 \le j \le L + n, \tag{4b}$$

$$Q^{n+1} = \sum_{j=0}^{L+n} b_j^{n+1} U_j^{n+1}, \tag{4c}$$

$$F_{2,j}^n = -f(a_{j+1}, Q^{n+1})(U_j^n + hF_{1,j}^n), \quad 0 \le j \le L + n, \tag{4d}$$

$$U_{j+1}^{n+1} = U_j^n + \frac{h}{2}(F_{1,j}^n + F_{2,j}^n), \quad 0 \le j \le L+n, \tag{4e}$$

$$\hat{Q}^{n+1} = \sum_{j=0}^{L+n+1} b_j^{n+1} \beta(a_j, Q^{n+1}) U_j^{n+1}, \tag{4f}$$

$$U_0^{n+1} = g(t^{n+1}, \hat{Q}^{n+1}), \tag{4g}$$

where the quadrature weights b_j^n are give by

$$b_j^n = \begin{cases} \dfrac{h}{3} & \text{if } j = 0 \text{ or } j = L+n; \\ \dfrac{2h}{3} & \text{if } j \text{ is odd;} \\ \dfrac{4h}{3} & \text{otherwise.} \end{cases}$$

Note that formulas (4a) and (4b) are explicit, while all other formulas are implicit. In practice system (4c)-(4g) is solved by successive iterations. In the next section we show that in an appropriately chosen space the map defined by (4c)-(4g) is a contraction. This guarantees that the iterations converge and the whole process (4) is well defined.

Regularity conditions. To conclude this section we mention that convergence analysis of any timestepping algorithm and any quadrature formula require some regularity of $u(t, a)$ and functions f, g, β. In the sequel we assume that the following holds:

(H_1)

$u_0(a)$ is bounded, nonnegative and there exists a maximum value A such that $u_0(a) = 0$ for $a > A$;

(H_2)

$f, \beta \in C^3([0, \infty] \times [0, +\infty))$, $\beta(\cdot, P)$, $f(\cdot, P)$, $\beta_P(\cdot, P)$, $f_P(\cdot, P) \in C^1([0, \infty), L_\infty[0, \infty))$, $\beta(\cdot, P)$, $f(\cdot, P)$ are compactly supported in $[0, \infty)$, there exists a positive constant C such that $0 \le \beta(a, P) < C$;

(H_3)

$g \in C^1([0, T] \times \mathcal{D}_2)$, where $\mathcal{D}_2$ is a compact neighborhood of

$$\left\{ \int_0^{+\infty} \beta(a, P(t)) u da,\ 0 \le t \le T \right\}.$$

(H_4)

$$u^0(0) = g(0, z^0),$$

where $z^0 = \int_0^{+\infty} \beta(a, P^0) u^0(a) da$ and $P^0 = \int_0^{\infty} u^0(a) da$.

(H_5)

$$u_a^0(0) = -[f(0, P^0) + \beta(0, P^0)] u^0(0) + g_t(0, z^0) - g_z(0, z^0)$$

$$\times \left[\int_0^{Runge-Kuttamethods,\ infty} \{\beta_a(a, P^0) + \beta_P(a, P^0) P_t^0 \right.$$

$$\left. - \beta(a, P^0) f(a, P^0)\} u^0 da \right],$$

where $P_t^0 = u^0(0) - \int_0^{\infty} f(a, P^0) u^0(a) da$.

Hypotheses (H_1)-(H_3) guarantee existence of unique solution $u(t, a)$, while compatibility conditions (H_4)-(H_5) ensure that the solution is at least one time continuously differentiable. We refer to [7, 8] where the following theorem is proven:

Theorem 2.1. *Assume that* (H_1)-(H_3) *hold, then the problem* (1) *has unique non negative solution* $u(t, a)$, *global in time and* $u \in C^1([0, T] \times [0, \infty) / \{(t, t) | t \ge 0\})$. *If in addition* (H_4)-(H_5) *are satisfied then* $u \in C^1([0, T] \times [0, +\infty))$.

To ensure that $u \in C^k([0, T] \times [0, +\infty))$, $k \ge 2$, in addition to (H_1)-(H_5) one has to assume

(H_6^k)

$$\lim_{a\to 0}\frac{d^l}{da^l}u^0(a)=\lim_{t\to 0}\frac{d^l}{dt^l}g(t,\,P(t)),\quad 2\le l\le k.$$

Compatibility condition (H_6^k) plays essential role in the proof of the consistency of the scheme.

3. Convergence Analysis

To simplify the analysis we introduce some notation. Assume that time step h takes values in $H=\{T/N,\ n\in\mathbb{N},\ N>0\}$, for a given n and h we define the normed space

$$\mathbb{X}^n=\mathbb{R}^{L+n},\quad \|U^n\|_n=h\sum_{j=0}^{L+n}|U_j^n|,$$

the map

$$\Psi_h^n:\mathbb{X}^n\times\mathbb{X}^{n+1},\quad \Psi_h^n(U^n,U^{n+1})=\left(\frac{1}{h}g(t^{n+1},\hat{Q}^{n+1}),\frac{1}{2}(F_1^n+F_2^n)\right)^T,$$

and the matrix

$$J^n=\begin{bmatrix}0\\ \\ I\end{bmatrix},\quad I\in\mathbb{R}^{(L+n)\times(L+n)}.$$

Then method (4) takes the form

$$U^{n+1}=J^nU^n+h\psi_h^n(U^n,U^{n+1}),\quad 0\le n\le N-1. \tag{5}$$

We set

$$U_h^n=(u(t^n,a_0),...,u(t^n,a_{L+n}))^T\in\mathbb{X}^n,$$

and assume that $B_R(U)\subset\mathbb{X}^n$ denotes the ball of radius R centered at U.

First, we show that Ψ_h is Lipschitz continuous.

Lemma 3.1. *Assume that hypotheses* (H_1)-(H_4) *hold and*

$$V^n, W^n \in B_R(U_h^n) \subset \mathbb{X}^n \text{ and } V^{n+1}, W^{n+1} \in B_R(U_h^{n+1}) \subset \mathbb{X}^{n+1},$$

for some $R > 0$. *Then*

$$\| \Psi_h^n(V^n, V^{n+1}) - \Psi_h^n(W^n, W^{n+1}) \|_{n+1}$$

$$\geq L_\Psi(\| V^n - W^n \|_n + \| V^{n+1} - W^{n+1} \|_{n+1}), \tag{6}$$

where L_Ψ *depends on* $L, T, R, U_h^n, U_h^{n+1}$ *and Lipschitz constants of* f, g, β.

Proof. Let Q_{V^n} and Q_{W^n} be quadrature rules for the components V^n and W^n, respectively. Then

$$| Q_{V^n} - Q_{W^n} | \leq \sum_{j=0}^{L+n} b_j^n | V_j^n - W_j^n | \leq \frac{4}{3} \| V^n - W^n \|_n.$$

Similarly, if $\hat{Q}_{V^n}$ and $\hat{Q}_{W^n}$ are quadrature approximations of $\int_0^\infty \beta(a, P(t^n))$, then (H_2) and previous estimate imply

$$| \hat{Q}_{V^n} - \hat{Q}_{W^n} | \leq \frac{4}{3} h \sum_{j=0}^{L+n} | \beta(a_j, Q_{V^n}) - \beta(a_j, Q_{W^n}) |$$

$$\leq \frac{4}{3} h \sum_{j=0}^{L+n} L_\beta | Q_{V^n} - Q_{W^n} |$$

$$\leq \frac{16}{9} L_\beta (L + T) \| V^n - W^n \|_n,$$

where L_β is the Lipschitz constant of β.

Next, we estimate the components of Ψ_h^n. For the first component, we have

$$\frac{1}{h} | (t^{n+1}, \hat{Q}_{V^{n+1}}) - g(t^{n+1}, \hat{Q}_{W^{n+1}}) | \leq \frac{1}{h} L_g | \hat{Q}_{V^{n+1}} - \hat{Q}_{W^{n+1}} |$$

$$\leq \frac{16}{9h} L_g L_\beta (L+T) \| V^{n+1} - W^{n+1} \|_{n+1},$$

where L_g is the Lipschitz constant of g. Lengthy but straightforward calculations show that all other components of

$$E_j^n = \Psi_h^n(V^n, V^{n+1}) - \Psi_h^n(W^n, W^{n+1}), \quad 1 \leq j \leq L+n+1,$$

are bounded by

$$| E_j^n | \leq (1+h) M | V_j^n - W_j^n | + (1+h) B L_f \| V^n - W^n \|_n$$

$$+ \frac{4}{3} (B + hM) L_f \| V^{n+1} - W^{n+1} \|_{n+1},$$

where L_f is the Lipschitz constant of f,

$$B \leq \max\{\| U_h^n \|_n, \| U_h^{n+1} \|_{n+1}\} + R \text{ and } M \leq \max_{(a,t) \in \mathbb{R}_+ \times [0,T]} | f(a, u(t, a)) |.$$

Putting all together we obtain (6). □

Using Lemma 3.1 it is not difficult to prove that discretization (5) is stable.

Lemma 3.2. *Let* V^n, W^n *be two sequences such, that* V^n, $W^n \in B_R(U_h^n)$, *for* $0 \leq n \leq N$. *Then for sufficiently small* $h \in H$ *the following estimate holds*

$$\| V^{n+1} - W^{n+1} \|_{n+1} \leq \frac{S}{h} \max_{0 \leq j \leq n+1} \| R^j \|_j, \tag{7}$$

where

$$R^{n+1} = V^{n+1} - W^{n+1} - J^n (V^n - W^n)$$

$$- h(\Psi_h^n(V^n, V^{n+1}) - \Psi_h^n(W^n, W^{N+1})), \quad 0 \leq n \leq N-1$$

$$R^0 = h(V^0 - W^0),$$

and S does not depend on h.

Proof. We denote $\delta^n = \| V^n - W^n \|_n$. Using Lemma 3.1 and the identity

$$V^{n+1} - W^{n+1}$$

$$= J^n(V^n - W^n) + h(\Psi_h^n(V^n, V^{n+1}) - \Psi_h^n(W^n, W^{N+1})) + R^{n+1},$$

we infer

$$(1 - hL_\Psi)\delta^{n+1} \le (1 - hL_\Psi)\delta^n + \| R^{n+1} \|_{n+1}.$$

Straightforward application of discrete Gronwall's inequality yields

$$\delta^{n+1} \le \left(\frac{1 + hL_\Psi}{1 - hL_\Psi}\right)^{n+1} \| R^0 \|_0 + \frac{1}{1 - hL_\Psi} \sum_{j=0}^{n} \left(\frac{1 + hL_\Psi}{1 - hL_\Psi}\right)^{n+1-j} \| R^{j+1} \|_{j+1}$$

$$\le \frac{2}{hL_\Psi(1 - hL_\Psi)} \left(\frac{1 + hL_\Psi}{1 - hL_\Psi}\right)^{n+1} \max_{0 \le j \le n+1} \| R^j \|_j.$$

Since $\frac{2}{L_\Psi(1 - hL_\Psi)} \left(\frac{1 + hL_\Psi}{1 - hL_\Psi}\right)^{n+1}$ is uniformly bounded for all $0 \le n \le N$, estimate (7) follows. □

Next, we show that method (5) is consistent, provided that (H_6^2) holds.

Lemma 3.3. *Assume that hypotheses* (H_1)-(H_5) *and* (H_6^k), $k \ge 2$, *are satisfied, then*

$$\| U_h^{n+1} - J^n U_h^n - h\Psi_h^n(U_h^n, U_h^{n+1}) \|_{n+1} = \mathcal{O}(h^{\min\{k-1, 3\}}), \quad 0 \le n \le N - 1. \quad (8)$$

Proof. Hypotheses (H_1)-(H_5) and (H_6^2), $k \ge 2$, imply that $u \in C^k([0, T) \times [0, \infty))$. Let

$$\overline{U}^{n+1} = J^n U_h^n + h\Psi_h^n(U_h^n, U_h^{n+1}),$$

then, taking into account orders of Heun's method (the order is 2) and composite Simpson's rule (the order is 4), we infer

$$Q^n = P^n + \mathcal{O}(h^{\min\{k, 4\}}), \quad (9a)$$

$$F_{1,j}^{n} = -f(a_j, P^n)U_{h,j}^{n} + \mathcal{O}(h^{\min\{k,4\}}), \quad 0 \le j \le L+n, \tag{9b}$$

$$Q^{n+1} = P^{n+1} + \mathcal{O}(h^{\min\{k,4\}}), \tag{9c}$$

$$F_{2,j}^{n} = -f(a_{j+1}, P^{n+1})U_{h,j+1}^{n} + \mathcal{O}(h^{\min\{k,4\}}), \quad 0 \le j \le L+n, \tag{9d}$$

$$\overline{U}_{j+1}^{n+1} = U_{h,j}^{n} + \frac{h}{2}(F_{1,j}^{n}, F_{2,j}^{n}) = U_{h,j+1}^{n+1} + \mathcal{O}(h^{\min\{k,3\}}), \quad 0 \le j \le L+n, \tag{9e}$$

$$\hat{Q}^{n+1} = \int_0^{\infty} \beta(a, P^{n+1})u(a, t^{n+1})da + \mathcal{O}(h^{\min\{k-1,3\}}), \tag{9f}$$

$$\overline{U}_0^{n+1} = g(t^{n+1}, Q^{n+1}) = U_{h,0}^{n+1} + \mathcal{O}(h^{\min\{k-1,3\}}), \tag{9g}$$

or shortly

$$\overline{U}^{n+1} = U_h^{n+1} + \mathcal{O}(h^{\min\{k-1,3\}}),$$

and (8) follows. □

We conclude this section with the following theorem.

Theorem 3.4. *Assume that hypotheses* (H_1)-(H_5) *and* (H_6^k) *with* $k \ge 3$, *are satisfied. If* $U^0 \in B_R(u_h)$, *then for all sufficiently small* $h \in H$ *the following is true*:

(*a*) *for each* $1 \le n \le N$ *there exists a unique solution of* (5) $U^n \in B_R(U_h^n)$;

(*b*) *the numerical solution converges and*

$$\| U_h^n - U^n \|_n \le \mathcal{O}(\| U^0 - U_h^0 \|_0 + h^{\min\{k-2,2\}}), \quad 0 \le n \le N. \tag{10}$$

Proof. We use induction. Obviously theorem holds for $n = 0$. Assume that statements (a), (b) are true for $0 \le k \le n$. We show then that it holds for $k = n+1$.

Part (a). Consider the iterations:

$$U^{n+1,i+1} = J^n U^n + h\Psi_h^n(U^n, U^{n+1,i}), \quad i \ge 0,$$

$$U^{n+1,0} = (U_0^n, U_0^n, ..., U_{L+1}^n).$$

First of all, induction hypothesis imply $U^{n+1,0} \in B_{\mathcal{O}(h^{\min\{k-2,2\}})}(U_h^n)$. Second, using the regularity assumptions we infer $U_h^n \in B_{\mathcal{O}(h)}(U_h^{n+1})$. Hence,

$$U^{n+1,0} \in B_{\mathcal{O}(h^{\min\{k-2,2\}})}(U_h^n) \subset B_{\mathcal{O}(h)}(U_h^{n+1}) \subset B_R(U_h^{n+1}),$$

provided that h is small.

Next, let $e_i = \| U^{n+1,i+1} - U^{n+1,i} \|_{n+1}$, then for the first iterate we have

$$e_1 \leq | U_0^n | + hM \leq h(M + \| U^{n+1,0} \|_{n+1}).$$

The inequality above implies that $U^{n+1,1} \in B_{\mathcal{O}(h)}(U^{n+1,0})$ and we conclude that $U^{n+1,1} \in B_R(U_h^{n+1})$, provided that h is small.

In general Lemma 3.1 implies

$$e_i \leq (hL)e_{i-1} \leq (hL_\Psi)^{i-1}h(M + \| U^{n+1,0} \|_{n+1}),$$

therefore

$$\| U^{n+1,i+1} - U^{n+1,0} \|_{n+1} \leq h(M + \| U_0^{n+1} \|_{n+1}) \sum_{j=0}^{i-1} (hL_\Psi)^j$$

$$\leq h \frac{M + \| U_0^{n+1} \|_{n+1}}{1 - hL_\Psi}.$$

We conclude that the iterates do not leave the ball $B_{\mathcal{O}(h)}(U^{n+1,0}) \subset B_R(U_h^{n+1})$, and converge to unique $U^{n+1} \in B_{\mathcal{O}(h)}(U^{n+1,0}) \subset B_R(U_h^{n+1})$, that solves (5).

Part (b). Combining Lemmas 3.2 and 3.3, we obtain

$$\| U^{n+1} - U_h^{n+1} \|_{n+1}$$

$$\leq \frac{S}{h} (\| U^0 - U_h^0 \|_0 + \max_{0 \leq i \leq n} \| U_h^{i+1} - J^i U^i - h\Psi_n^i(U_h^i, U_h^{i+1}) \|_{i+1})$$

$$\leq \frac{S}{h} (h\| U^0 - U_h^0 \|_0 + \mathcal{O}(h^{\min\{k-1,3\}}))$$

$$= \mathcal{O}(\| U^0 - U_h^0 \|_0 + h^{\min\{k-2,\,2\}}).$$

The proof is complete. □

4. Numerical Results

In this section, we provide three computational examples. In the first one, we study efficiency of (4) in term of global error, CPU time and order of convergence. Note that in this example the maximum age is finite. In the second example we compare numerical solutions obtained by (4) and by the low order explicit Euler scheme (once again the maximum age is set to be finite). The last example deals with the case of infinite maximum age. In all the examples the global error is calculated by means of the formula

$$E_h = \max_{0\le n\le N} \| U_h^n - U^n \|_n \tag{11}$$

Example 1. In this example the maximum age A is set to be 1, the age-specific fertility and mortality moduli, the initial density and the birth function are given by

$$f(a,\, z) = -z, \quad \beta(a,\, z) = \frac{aze^{-a}}{(1+z)^2}, \quad u^0(a) = \frac{e^{-a}}{2 - e^{-A}},$$

$$g(z,\, t) = \frac{4z(2 - 2e^{-A} + e^{-t})^2}{(1 - e^{-A})\,(1 - (1+2A)e^{-2A})\,(2 - 2e^{-A} + e^{-t})},$$

respectively. The exact solution is (see [2])

$$u(t,\, a) = \frac{e^{-a}}{1 - e^{-A} + e^{-t}}.$$

Results of numerical simulations (in time interval [0, 10]) are shown in Table 1. Thanks to the choice of time stepping algorithm (there is no off-grid stage values) very accurate approximations are obtained in a reasonably short time. Note also that $u(t,\, a) \in C^\infty([0,\, T) \times [0,\, 1))$ and Theorem 3.4 predicts that $E_h = \mathcal{O}(h^2)$. The last row of Table 1 is in complete agreement with the theory. The exact and the numerical solutions and the pointwise errors are shown in Figure 1.

Table 1. Errors, CPU time (seconds), and order of convergence Example 1.

N	20	40	80	160	320
E_h	$0.534 \cdot 10^{-5}$	$0.212 \cdot 10^{-5}$	$0.325 \cdot 10^{-6}$	$0.796 \cdot 10^{-7}$	$0.195 \cdot 10^{-7}$
CPU	0.094	0.187	0.515	1.497	6.74
order		2.049	1.988	2.030	2.027

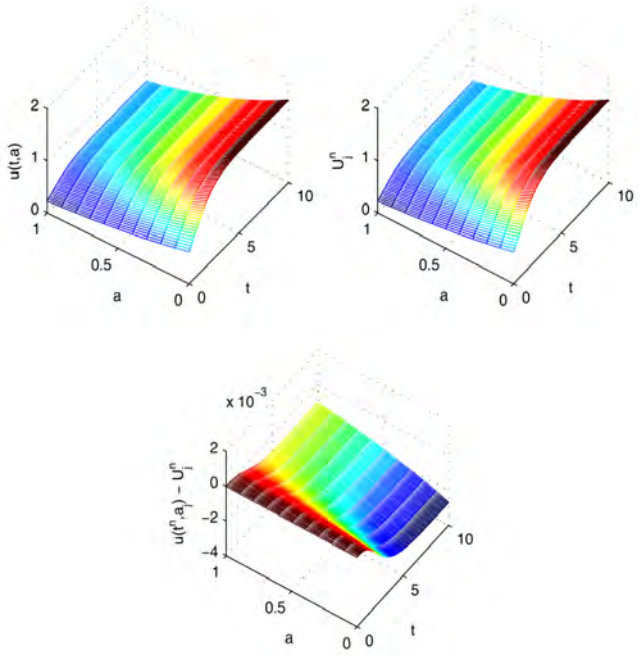


Figure 1. The exact solution, the numerical solution and the pointwise error, $N = 100$.

Example 2. Here we compare numerical solutions obtained using (4) and using the low order explicit Euler scheme. We use the following data: $A = 0.9$, $T = 1$,

$$f(a, z) = -\frac{1}{1-a} - z, \quad \beta(a, z) = 4, \quad g(t, z) = z, \quad u^0(a) = 4(1-a)e^{-\alpha a}.$$

The exact solution is given by

$$u(a, t) = \frac{4\alpha(1-a)e^{-\alpha a}}{(\alpha-1)e^{-\alpha t}+1},$$

where $\alpha = 2.5569290855$.

The numerical solutions obtained by explicit Euler method and by (4) are shown in Figure 2. Note that the low order solutions develop jump discontinuity along the main characteristic line $a = t$. The magnitude of the jump decreases as $N \to \infty$ but does not vanish completely. The situation is totally different if we use (4), the numerical solutions remain smooth for all N.

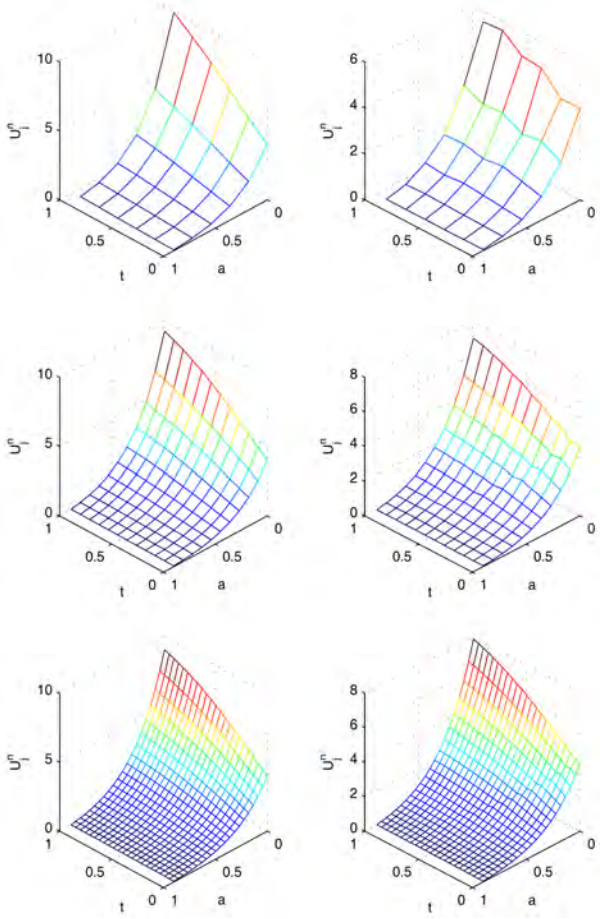


Figure 2. The numerical solution obtained by (4) and by explicit Euler Scheme (left and right columns, respectively), $N = 5, 10, 20$.

Example 3. In this example $A = \infty$, the age-specific fertility and mortality moduli are the same as in Example 1, the birth function and the initial age-specific density are given by

$$g(t, z) = \frac{4z(2 + e^{-t})^2}{(1 + e^{-t})}, \quad u^0(a) = \frac{1}{2} e^{-a},$$

respectively. The exact solution is

$$u(t, a) = \frac{e^{-a}}{1 + e^{-t}}.$$

For numerical simulations we took L to be large $(5N \le L \le 2000N)$. It is clear that the cost of computations increases together with L. However, the theory developed in Section 3 asserts that $E_h = \mathcal{O}(h^{\min\{k-1, 2\}})$ as soon as supp $u^0 \subset [0, hL]$. Practical implication | there is no need to increase L indefinitely, the accuracy of computations will be the same for all $L \ge L^*$, where L^* is the smallest positive integer that satisfy supp $u^0 \subset [0, hL^*]$. Results of simulations (in time interval $[0, 10]$) shown in Table 2 completely confirms this simple observation.

5. Conclusion

The numerical scheme developed in this paper provides a good approximation to the age structured population model. Analysis shows that the convergence rate is $\mathcal{O}(h^2)$, provided that the exact solution is sufficiently smooth. Moreover, the order of convergence is independent of the age interval. If the maximum age is not finite it is sufficient to choose parameter L so, that supp $u^0 \subset [0, hL]$.

To conclude, we mention that analysis of Section 3 requires global regularity of $u(t, a)$ in $[0, T) \times [0, \infty)$. However, the method can be reformulated to cope with discontinuities along characteristic lines. In addition, the method can be adapted to size-structured models [9]. In this case, the characteristic curves are not straight lines and more complicated meshing strategy must be used.

Table 2. Errors and CPU time (seconds) Example 3.

L	N	20	40	60	80
$L = 5N$	E_h	$0.121 \cdot 10^{-1}$	$0.233 \cdot 10^{-1}$	$0.36 \cdot 10^{-1}$	$0.482 \cdot 10^{-1}$
	CPU	0.109	0.20	0.327	0.499
$L = 10N$	E_h	$0.816 \cdot 10^{-5}$	$0.421 \cdot 10^{-5}$	$0.306 \cdot 10^{-5}$	$0.124 \cdot 10^{-5}$
	CPU	0.14	0.29	0.515	0.827
$L = 1000N$	E_h	$0.782 \cdot 10^{-5}$	$0.335 \cdot 10^{-5}$	$0.294 \cdot 10^{-5}$	$0.169 \cdot 10^{-5}$
	CPU	3.1	10.28	22.94	42.073
$L = 2000N$	E_h	$0.782 \cdot 10^{-5}$	$0.335 \cdot 10^{-5}$	$0.294 \cdot 10^{-5}$	$0.169 \cdot 10^{-5}$
	CPU	4.618	17.082	38.204	70.184

References

[1] L. M. Abia, O. Angulo and J. C. Lòpez-Marcos, Age-structured population dynamics models and their numerical solutions, Ecol. Model. 188 (2005), 112-136.

[2] L. M. Abia and J. C. Lòpez-Marcos, Runge-Kutta methods for age-structured population models, Appl. Num. Math. 17 (1995), 1-17.

[3] O. Angulo, J. C. Lòpez-Marcos, M. A. Lòpez-Marcos and F. A Milner, A numerical method for nonlinear age-structured population models with finite maximum age, J. Math. Anal. Appl. 361 (2010), 150-160.

[4] M. Iannelli, Mathematical Theory of Age-Structured Population Dynamics, Appl. Math. Monographs, C.N.R., Giardini Editori e Stampatori, Pisa, 1994.

[5] M. Iannelli, M. Martcheva and F. A. Milner, Gender-Structured Population Modeling: Mathematical Methods, Numerics and Simulations, SIAM, Philadelphia, 2005.

[6] J. C. Lòpez-Marcos and J. M. Sanz-Serna, Stability and convergence in numerical analysis III: Linear investigation of nonlinear stability, J. Num. Anal. 8 (1988), 71-84.

[7] M.-Y. Kim, Numerical methods for a stiff problem arising from population dynamic, Kangweon-Kyungki Math. J. 13(2) (2005), 161-176.

[8] M. E. Gurtin and R. C. MacCamy, Nonlinear age-dependent population dynamics, Arch. Ration. Mech. Anal. 54 (1974), 281-300.

[9] K. Tanya, An explicit third-order numerical method for size-structured population equations, Num. Meth. PDE's 19(1) (2002), 1-21.

[10] G. F. Webb, Theory of Age Dependent Population Dynamics, Marcel Dekker, New York, 1985.

ANNEXE 4

Article :

H.NKOUNKOU1, A.TRAORE, M. S. D. HAGGAR and B. MAMPASSI, *Solving Convection Diffusion Problem With a Pseudo Spectral Method on Unstructured Meshes*, pioneer Journal of Computer Science and Engineering Technology Volume , N° 1-2, pp. 1--12 (2014)


Pioneer Journal of Computer Science and Engineering Technology
Volume 7, Numbers 1-2, 2014, Pages 1-12
This paper is available online at http://www.pspchv.com/content_PJCSET.html

SOLVING CONVECTION DIFFUSION PROBLEMS WITH A PSEUDO SPECTRAL METHOD ON UNSTRUCTURED MESHES

HILAIRE NKOUNKOU[1], ABOUBAKARI TRAORE[2], MAHAMAT SALEH DAOUSSA HAGGAR[3] and BENJAMIN MAMPASSI[4]

[1]Marien Ngouabi University
Congo
e-mail: hnkounkou@yahoo.fr

[2]Ecole Normale Supérieur d'Abidjan
Ivory Cost
e-mail: traboubakari@yahoo.fr

[3]N'Djamena University
Chad
e-mail: daoussa_haggar@yahoo.fr

[4]Dakar University
Senegal
e-mail: mampassi@yahoo.fr

Abstract

We saw in [H. Nkounkou, A. Traoré, G. Séworé, M. A. Abani and B. Mampassi, Least squares collocation methods for solving partial differential equations: a matlab approach, Pioneer Journal of Computer Science and Engineering Technology 1(2) (2011), 57-71; H. Nkounkou, A. Traore, G. Seworé, A. M. Abani and B. Mampassi, Spectral differentiation on unstructured meshes using Jacobi Gauss-Lobatto points, Far East Journal of Applied Mathematics 59(2) (2011), 105-122] that the

Received September 30, 2014

Keywords and phrases: Gauss-Lobatto points, differentiation matrices, triangular meshes, least squares collocation methods, complex geometry domains.

approximation of differentiation operators of order 2 was less good because of errors in the second derivation. To compensate this problem, we propose in this article an alternative method for approximating differentiation operators of order 2 by means of collocation points and we present an algorithm for solving a convection diffusion problem in complex geometry domains using differentiation matrices with Jacobi Gauss Lobatto points. We develop Matlab codes for solving the problem by a Least squares collocation method. Least squares collocation methods are considered as alternative to least squares finite elements methods [W. Heinrichs, An adaptive spectral least squares spectral collocation method with triangular elements for the incompressible Navier-Stokes equations, Springer and Business Media (2006), 337-350; D. Pathria and G. E. Karniadakis, Spectral element methods for elliptic problems in non smooth domains, Journal of Computational Physics 122 (1995), 83-95] and are particularly very attractive for solving partial differential equations on complex geometry domains. A numerical example is supporting our paper.

1. Introduction

Most of the physical problems are governed by non-linear partial differential equations whose analytical resolution is very difficult or impossible in many cases. Hence the need of using approximation methods based generally on differentiation matrices [7]. The choice of Gauss-Lobatto collocation points in the pseudo spectral approximation is an important factor for triangular elements [1, 6]. It is well known that computing solutions of Partial Differential Equations (PDEs) by a Least Squares Collocation method [2], requires accurate approximation schemes for differential operators and occurs in three stages:

- The generation of the mesh of the studied domain as well as corresponding collocation points;
- The discretization of the problem on each element by the spectral method, followed by the process of assembly of local solutions;
- The determination of the overall solution using approximating differentiation operators by means of collocation points.

The outline of this paper is as follows. In Section 2, we present the linear convection diffusion problem. In Section 3, we describe the discretisation of this problem, and give its formulations of differentiation matrices associated with

two-dimensional mesh domains. This section gives also the algorithm for solving the convection diffusion problem and Matlab codes for computing differentiation matrices [7]. In the last section, we present numerical experiments and we discuss on the accuracy of the method.

2. The Model Problem

As a model problem, we consider the following convection diffusion system

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega =]-1,1[^2 \setminus]0,1[^2, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1}$$

where $f \in L^2(\Omega)$. Referring to [3], the system (1) admits a non-trivial unique solution in $H_0^1(\Omega)$. We propose, here, an alternative approach of collocation methods to approximating second order differentiation operators. We saw in [9] that direct application of the collocation method to second order differential operators was less good due to errors proliferation in the second derivation. To overcome this problem, it is worthwhile to transform the system (1) by a system of partial differential equations of the first order. To this end, let introduce two new unknowns v and w by setting

$$(v,\ w)^T\ =\ \overrightarrow{grad}\ u.$$

It then follows:

$$\begin{cases} -div\begin{pmatrix} v \\ w \end{pmatrix} + u = f, \\ \begin{pmatrix} v \\ w \end{pmatrix} = \overrightarrow{grad}\, u, \end{cases} \tag{2}$$

with boundary conditions $u = 0,\ \ v = 0$ and $w = 0$ on $\partial\Omega.$. Explicitly, we have to seek $(u,\ v,\ w)$ solution of the problem:

$$\begin{cases} \dfrac{\partial u}{\partial x} - v = 0 & \text{in } \Omega, \\ \dfrac{\partial u}{\partial y} - w = 0 & \text{in } \Omega, \\ -\dfrac{\partial v}{\partial x} - \dfrac{\partial w}{\partial x} + u = f & \text{in } \Omega, \\ u = v = w = 0 & \text{on } \partial\Omega. \end{cases} \tag{3}$$

This system of partial differential equations of the first order has some advantage for approximating derivatives with the pseudo spectral method. This latter system is therefore to be considered in the rest of this paper for developing a numerical scheme based on the combination of both collocation and least squares methods which integrates the complexity of the domain geometry.

3. Collocation Least Squares Approximation

Let U, V, W be the vectors whose components are, respectively the values of u, v, w at collocation points, associated with elements of the meshes $\mathcal{T}_h$ as defined in [9]. Considering differentiation matrices described in [9] the discretization of the problem (3) leads to:

$$\begin{cases} D^x U - V = 0, \\ D^y U - W = 0, \\ -D^x V - D^y W + U = F, \\ D_b U = 0, \\ D_b V = 0, \\ D_b W = 0, \end{cases} \tag{4}$$

where D^x and D^y are pseudo differential matrices for which if U_x and U_y denote vectors of values of derivatives respect to x and y at collocation points, then we have $U_x = D^x U$ and $U_y = D^y U$. Here, D_b is the boundary matrix, F the vector of values of f at collocation points. It should be noticed that the accuracy of pseudo differentiation approximation has been studied in [8] and [9]. It is shown that this approximation is very attractive. The discretization problem (4) can be rewritten into block matrices forms.

$$\begin{pmatrix} D^x & -I_d & \underline{0} \\ D^y & \underline{0} & -I_d \\ I_d & -D^x & -D^y \\ D_b & \underline{0} & \underline{0} \\ \underline{0} & D_b & \underline{0} \\ \underline{0} & \underline{0} & D_b \end{pmatrix} \times \begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} \underline{0} \\ \underline{0} \\ F \\ \underline{0} \\ \underline{0} \\ \underline{0} \end{pmatrix}, \tag{5}$$

where I_d is the identity matrix. Setting

$$\mathcal{A} = \begin{pmatrix} D^x & -I_d & \underline{0} \\ D^y & \underline{0} & -I_d \\ I_d & -D^x & -D^y \\ D_b & \underline{0} & \underline{0} \\ \underline{0} & D_b & \underline{0} \\ \underline{0} & \underline{0} & D_b \end{pmatrix} \tag{6}$$

$X = (U, V, W)^T$ and $b = (\underline{0}, \underline{0}, F, \underline{0}, \underline{0}, \underline{0})^T$, we are then led in solving the linear system:

$$\mathcal{A}X = b. \tag{7}$$

The approximation of the problem (1) by the least squares collocation method is then reduced in finding X^* solution of

$$\| b - \mathcal{A}(X^*) \|^2 = \min \| b - \mathcal{A}(X) \|^2. \tag{8}$$

The resolution of such a problem can then be obtained in MatLab by the simple command line:

$$X = \mathcal{A} \backslash b. \tag{9}$$

We know that there are matrices called assembling matrices Z and W [9] such that

$$D^x = \mathbf{W} D_x^{loc} \mathbf{Z} \tag{10}$$

and

$$D^y = \mathbf{W} D_y^{loc} \mathbf{Z}. \tag{11}$$

The differentiation matrices D^x and D^y, in the block matrix $\mathcal{A}$, are easily computed [9]. The boundary matrix D_d is constructed from knowledge of the mesh.

The differentiation global matrix [9] in which we define a third column to identify the edge points and to take into account the peaks of the domain. The code to compute the global points and the assembling matrices is done as follows:

```
%=================================================================
% MatLab code for computing the assembling matrices Z, W and Eglob
% (p,t) triangular refine of the domain by delaunay
% Eloc is the set of the local points of the refine triangles
% N the degree of the local interpolation polynomial
% Q generates the Gauss-lobatto points in a reference triangle
% (a,a) is the order of the Jacobi polynomial
function [Eglob,z,w]=pointscolocaux(p,t,N,a)
Eloc=numlocal(p,t,N,a);
E=round(Eloc(:,1:2)*1e4);
[E,I,J]=unique(E,'rows');
Eglob=Eloc(I,:);
for i=1:size(Eloc,1)
   for j=1:size(Eglob,1)
   if j==J(i)
        z(i,j)=1;
    end
   end
end
for i=1:size(Eglob,1)
   for j=1:size(Eloc,1)
   if j==I(i)
        w(i,j)=1;
    end
   end
end
%-----------------------------------------------------------------
%Indexing of boundary points
%-----------------------------------------------------------------
e = boundedges ( p, t ); %Computation of edges
SBORD=p(e(:,1),:);   %Vertices of the edge
E=round(Eglob(:,1:2)*1e4);SBORD=round(SBORD*1e4);
ISBORD=ismember(E,SBORD,'rows'); %index in Eglob of boundary points
%determination of index points of multiplicity 1 (indoor or on a
boundery edge)
for i=1:length(J)
    if length(find(J==J(i)))==1
        UNPOINT(J(i))=1;
    else
        UNPOINT(J(i))=0;
    end
```

```
end
% Generation of indexes of boundary points
for i=1:size(Eglob,1)
   if Eglob(i,3)==1 & UNPOINT(i)==1
       Eglob(i,3)=1;
    else
       Eglob(i,3)=0;
    end

end
Eglob(find(ISBORD==1),3)=1;
%=====================================================================
```

The code for the computation of the boundary matrix is given by:

```
%=====================================================================
% MatLab code for computing the boundary matrix
function Mb=Matbord(Eglob)
i=1;
for j=1:size(Eglob,1)
    if Eglob(j,3)==1
        Mb(i,j)=1;
        i=i+1;
    else
        Mb(i,j)=0;
    end
end
%=====================================================================
```

The general algorithm for solving the problem (1) may be done as follows.

Algorithm 3.1 (algorithm of solving the problem (1) by the LSCM method)**.**

(i) Read the parameter *N*, degree of local interpolation polynomial.

(ii) Load the settings of the mesh (p, t).

(iii) Compute the vectors and matrices associated with the mesh

- *Z*, ***W***: Assembling matrices.
- $U_{glob} = \boldsymbol{W} U_{loc}$.
- D_x^{Loc}; D_x^{Loc} : Local differentiation matrices.
- D^x; D^y : Global differentiation matrices.

- D_b : Boundering matrix.
- Define the matrix $\mathcal{A}$.

(iv) Compute:

- $U_{bord} = D_{bord} \times U_{glob}$: vector of boundering points.

(v) Compute: $X = \mathcal{A} \backslash b$.

4. Numerical Experiments

To test our algorithm, we have to consider the model problem (1) with

$$f(x, y) = (\pi^2 + 1)\sin(\pi x)\sin(\pi y).$$

It is easy to verify that the exact solution of this problem is

$$u(x, y) = \sin(\pi x)\sin(\pi y).$$

We propose then to compare the exact solution with approached solutions based on the mesh parameters and those of the spectral approximation. For the resolution of the system (7), we can use the normal approximation which consists to reduce this equation to the following

$$\mathcal{A}^T \mathcal{A} X = \mathcal{A}^T b. \tag{12}$$

It shows that it is quite equivalent to solve

$$\| \mathcal{A} X^* - b \|^2 = \min \| \mathcal{A} X - b \|^2, \tag{13}$$

but in this case it is reduced to use an optimization code such that: “LSQLIN” of MatLab.

The simulation results are presented in Figures 1 and 2. In Figure 1, we fixed the value of the discretization parameter $N = 4$ and varied the size of the mesh: A_1 (14 triangles), B_1 (16 triangles) and C_1 (28 triangles) and then we observed their corresponding approached solutions A_2, B_2 and C_2. In Figure 2, we fixed the size of the mesh and varied the value of discretization parameter *N*. We note that for this case, the direct method gives a better approximation than the method using the “LSQLIN” function, which is shown in the two following errors tables: Table 1 and

Table 2. We first give the errors table for $N = 4$ both for the direct method and for the method using the "LSQLIN" function.

Table 1. Evolution of the errors norm for the direct method with $N = 4$.

Meshes	**Euclidean norm of the Error**
14 triangles	0.61319
16 triangles	0.84072
28 triangles	1.2803

Table 2. Evolution of the errors norm for the method using « LSQLIN » function

Meshes	**Euclidean norm of the Error**
14 triangles	0.92762
16 triangles	1.1954
28 triangles	1.6575

When N increases and when the mesh becomes increasingly thin, we have a degradation of the approximation. It is due to bad conditioned matrix $[\mathcal{A}; \mathcal{B}]$, where the condition number of this matrix is computed on the basis of the mesh for $N = 4$.

Table 3. Condition number of the matrix $[\mathcal{A}; \mathcal{B}]$

Meshes	14 Triangles	16 Triangles	28 Triangles
Condition number of $[\mathcal{A}; \mathcal{B}]$	792.1456	542.2327	14133.

The Figure 2 represents approached solutions on a mesh grid of 16 triangles by varying the parameter of discretization N. We have here the confirmation that the approximation becomes less good when N increases however for values relatively modest, $N = 3$ and $N = 4$, the approximation is perfect. This justifies that the method converges well and that deterioration obtained for large values of N is simply due to the condition number which has an impact on predefined MATLAB optimization codes.

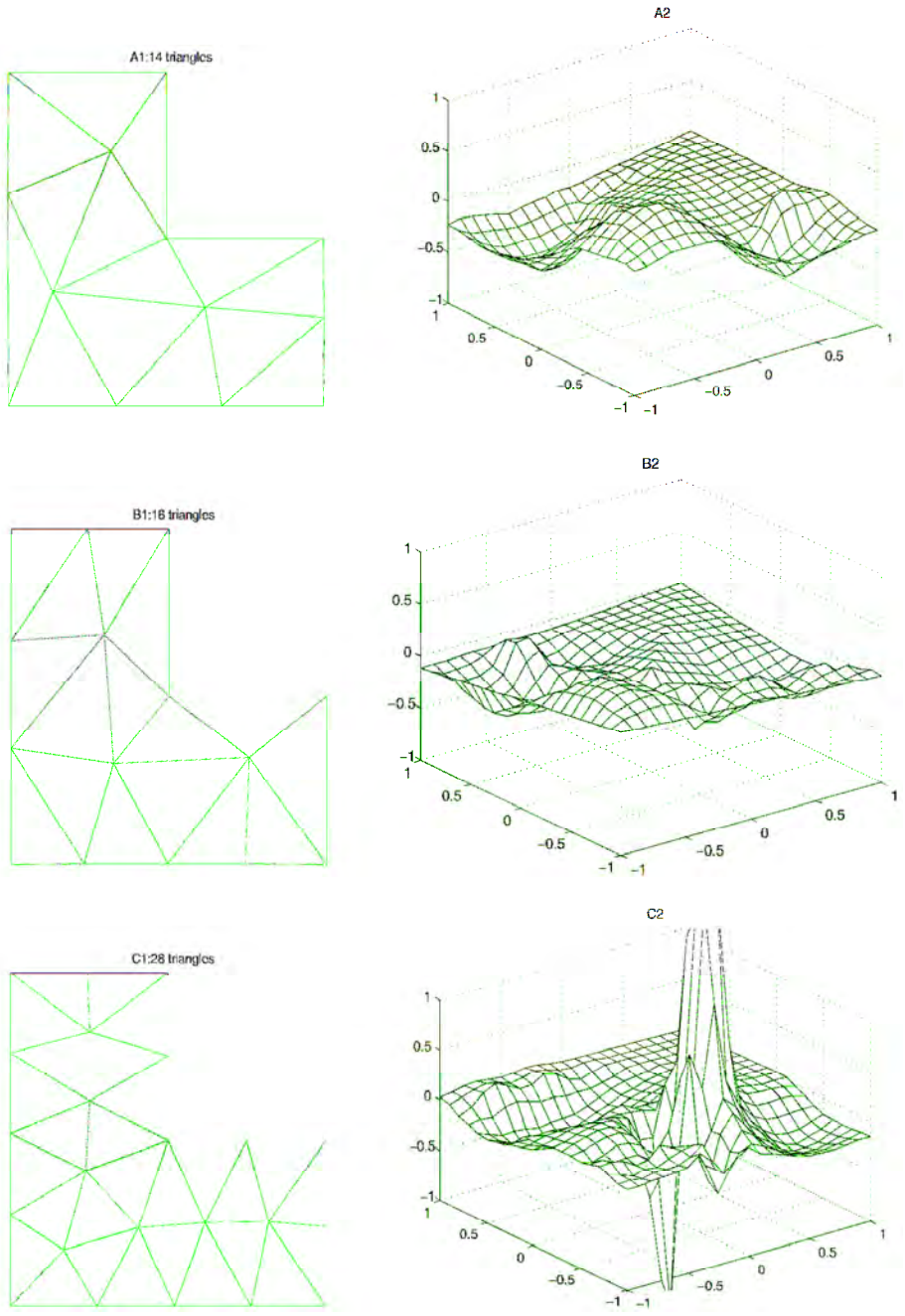


Figure 1. Meshes and approached solutions for $N = 4$.

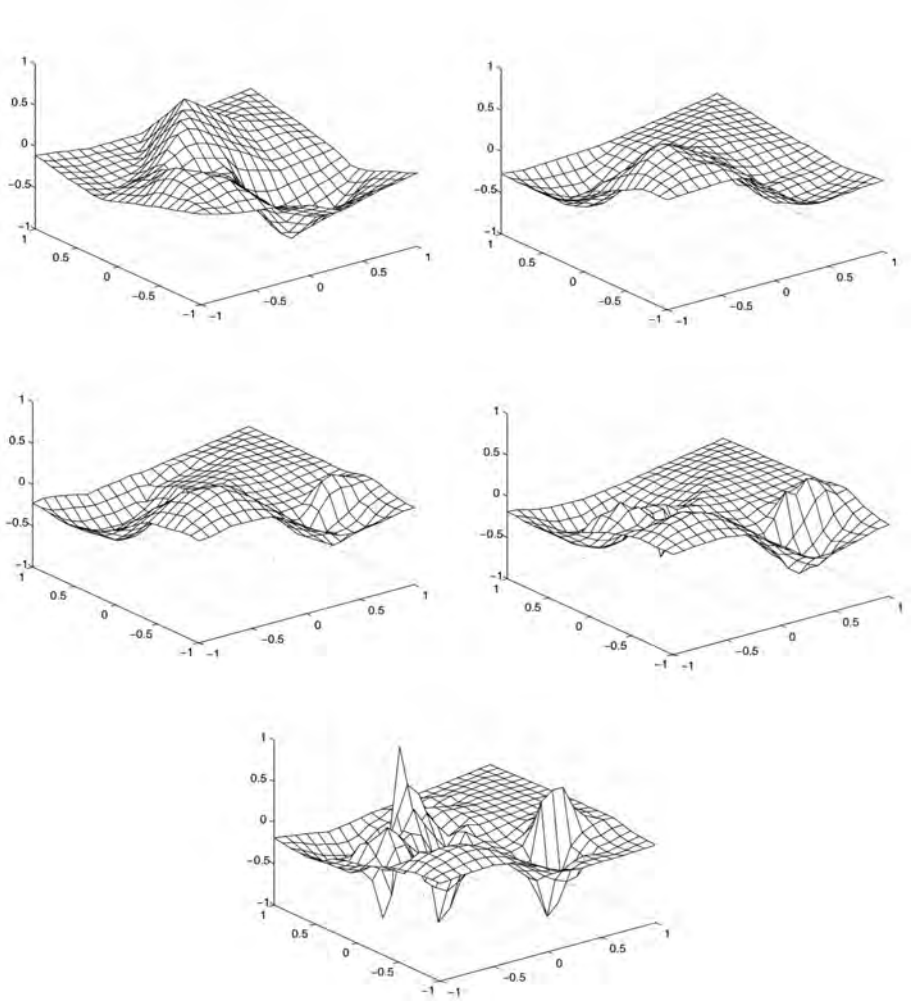


Figure 2. Approximated solutions for a fixed mesh within $N = 2, 3, 4, 5, 6$ respectively for figures from left to right.

References

[1] L. Bos, M. A. Taylor and B. A. Wingate, Tensor product Gauss-Lobatto points are Fekete points for the cube, Math. Comp. 70 (2001), 1543-1547.

[2] L. R. Bentley and G. F. Pinder, Solution of the advective-dispersive transport equation using a least squares collocation, Eulerian-Lagrangian method, Numer. Methods Partial Differential Equations 5(3) (1989), 227-240.

[3] H. Brezis, Analyse Fonctionnelle: Théorie et Applications, MASSON Paris, New York, Barcelone Milan Mexico, Sao Paulo, 1987.

[4] W. Heinrichs, An adaptive spectral least squares spectral collocation method with triangular elements for the incompressible Navier-Stokes equations, Springer and Business Media (2006), 337-350.

[5] D. Pathria and G. E. Karniadakis, Spectral element methods for elliptic problems in non smooth domains, Journal of Computational Physics 122 (1995), 83-95.

[6] Richard Pasquetti and Francesca Rapetti, Spectral element methods on triangles and quadrilaterals: comparisons and applications, Journal of Computational Physics 198 (2004), 349-362.

[7] L. N. Trefethen, Spectral methods in MATLAB, Software, Environments, and Tools, 10, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[8] H. Nkounkou, A. Traoré, G. Séworé, M. A. Abani and B. Mampassi, Least squares collocation methods for solving partial differential equations: a matlab approach, Pioneer Journal of Computer Science and Engineering Technology 1(2) (2011), 57-71.

[9] H. Nkounkou, A. Traore, G. Seworé, A. M. Abani and B. Mampassi, Spectral differentiation on unstructured meshes using Jacobi Gauss-Lobatto points, Far East Journal of Applied Mathematics 59(2) (2011), 105-122.

ANNEXE 5

Article :

N. KOSSADOUM, A. TRAORE, N. NGARKODJE, B. MAMPASSI, *On the numerical simulation of Lakes trying-up models*, Far East Journal of applied mathematics Volume79(2), pp. 111--126, (2013)

Far East Journal of Applied Mathematics
Volume 79, Number 2, 2013, Pages 111-126
Published Online: September 2013
Available online at http://pphmj.com/journals/fjam.htm
Published by Pushpa Publishing House, Allahabad, INDIA

ON THE NUMERICAL SIMULATION OF LAKES DRYING-UP MODELS

Kossadoum Ngarmadji[1], Aboubakari Traore[2], Ngarkodje Ngarasta[1] and Benjamin Mampassi[3]

[1]University of N'Djamena
Chad
e-mail: kossbri@yahoo.fr
ngarkodje@yahoo.fr

[2]Ecole Normale Superieure of Abidjan
Cote d'Ivoire
e-mail: traore_abou2001@yahoo.fr

[3]Cheikh Anta Diop University
Dakar, Senegal
e-mail: mampassi@yahoo.fr

Abstract

A partial differential equation (PDE) is proposed for simulating the lake drying-up. Precisely, we were interested in the change of the water level of Lake Chad in long-time. A couple of least squares and Donor-cell scheme are used to identify both the level and unknown parameters in a lake model. Numerical experiments give the behaviors of lake level and the values unknown parameters of model.

2010 Mathematics Subject Classification: 00A71, 35K55, 65L20.

Keywords and phrases: lake drying-up, Donor-cell scheme, least squares, lake level, satellite image.

Received May 8, 2013

1. Introduction

The shrinking and drying-up of Lake Chad in Africa have been widely observed in the recent years. In fact, Lake Chad was the sixth largest lake in the world 40 years ago but water level has decreased by more than 90 per cent of its area [2, 9, 4] (Figure 1). Political leaders, environmentalists and scientists have turned their attention to this dramatic topic and try to understand the causes, the water balance and how to stop the drying-up process.

Nowadays, many investigations are done in modeling and numerical analysis in order to build a useful tool to approximate the environmental problems by mathematical models and explain their behavior by numerical simulation. For this purpose, the aim of this paper is to contribute to the understanding of the drying-up of Lake Chad in long-time. Precisely, to develop a preliminary deterministic model of Lake Chad in order to estimate the water budget component of the lake and to simulate the lake level variations.

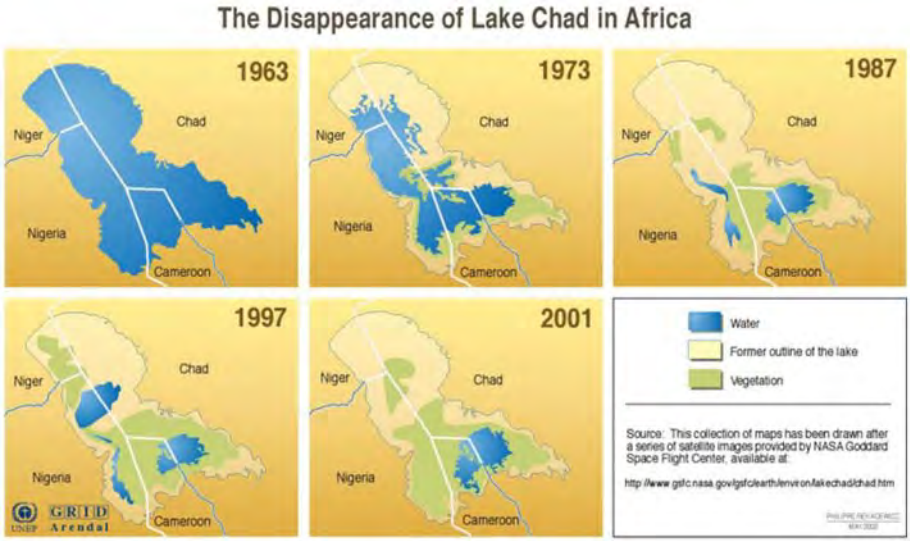


Figure 1. Chronology of Lake Chad change from 1963 to 2001. The collection of maps has been sourced from the series of satellite images provided by NASA Goddard Space Flight Center.

A similar investigation has been done in 2006 on Lake Tana in Ethiopia [8] where simulation of lake level variation has been conducted through modeling at a monthly time step.

The water balance studies of Africa lakes have been relatively well documented, the readers can see for instance, Lake Victoria [18, 14], Lake Malawi [11] and Lake Chad [5, 10, 13, 17]. The formers usually approach the problem by solving the differential water balance equation, simulating the lake level, integrating over a time interval and the differential equation can be rewritten as:

$$\frac{d}{dt}h(t) + E(t) = P(t) + \frac{V(t)}{A(h(t))} + \varepsilon(t), \tag{1}$$

where h is the lake level, A is the depth depending surface area of the lake, P is the rate of rainfall over the lake, E is the rate of the lake evaporation, t is the time and V is the volume of water exchanged by inflow, outflow and groundwater. The final term ε represents uncertainties in the water balance arising from errors in the data and other terms such as minor abstraction or inflow from ungauged catchments.

This model depends only on the variable time t, it means that the shape of the lake is neglected and the level h is supposed to be the same everywhere in the domain. It is clear that this is far to the reality if we consider the complexity of the climate, hydrological characteristics of the basins and the problem of sand at the bottom of the lake. The improvement in the existing water resources models or the development of news models required to take into account both the space and time variable which can be combined to produce sound models for lakes which generalizes the model given by (1).

This paper is organized as follows: in Section 2, we derive a mathematical model of drying-up process in a complex domain such as Lake Chad and we treat the discretization of the lake surface. In Section 3, a numerical scheme is proposed. Numerical experiments are presented in Section 4 and Section 5 is devoted to the conclusion.

2. Model

Lakes are complex dynamics systems due to the hydrology parameters (rain, evaporation, inflow and outflow) and physical characteristics (the basin, the boundary). These parameters can change from one lake to another. In this present paper, we focus our model on the physical characteristics of Lake Chad which can be adapted to other lakes. The model follows the physical laws and constitutive relations of modeling principles developed in [7] and the hydrobalance equation (1).

2.1. Derivation of equations

Let Ω be a bounded domain of $\mathbb{R}^2$ which represents the surface of lake, whose boundary is denoted by $\partial\Omega$ and assumed sufficiently regular. Moreover, let $h(\mathbf{x}, t)$ be the function which describes the level of the lake at position $\mathbf{x} \in \Omega$ and at time t. Let V be any material volume around $\mathbf{x}$ we deduce from (1) the water balance in X at time t by

$$\frac{d}{dt} h(\mathbf{x}, t) + E(\mathbf{x}, t) = P(\mathbf{x}, t) + \frac{V(\mathbf{x}, t)}{A(h)} + \varepsilon(t). \tag{2}$$

We assume that the hydrological parameters are not uniformly distributed so that the modeling parameters E, P also depend on the position variable. This assumption is realistic according to the surface of lake in 1963 (Figure 1). Let $\mathbf{q}(\mathbf{x}, t)$ be the flux of water exchanged such that the rate of water flow on a surface with unit normal $\mathbf{n}$ is

$$V(\mathbf{x}, t) = \mathbf{q}(\mathbf{x}, t) \cdot \mathbf{n} \tag{3}$$

per unit area $(A(h) = 1)$. Then, we can write down the conservation of mass for V in the form

$$\int_V \frac{d}{dt} h(\mathbf{x}, t) d\mathbf{x} + \int_V E(\mathbf{x}, t) d\mathbf{x} = \int_V P(\mathbf{x}, t) d\mathbf{x} + \int_{\partial V} \mathbf{q}(\mathbf{x}, t) \cdot \mathbf{n} d\mathbf{x} + \varepsilon(t), \tag{4}$$

where ∂V is the boundary of V.

We use Green's theorem on the surface integral and as V is arbitrary, equation (4) gives us

$$\frac{d}{dt}h + E = P + \nabla \cdot \mathbf{q} + \varepsilon(t). \tag{5}$$

At this point, we need to express the parameters E and $\mathbf{q}$ to the level h. We make the following assumption:

$$E = \alpha h^r, \quad 0 \le r \le 1, \tag{6}$$

where α and r are unknown real parameters and can be identified. The flux $\mathbf{q}$ is proportional to the volume hence to the level. The general expression of the flux is given by the transport equation and at this point, the assumption on the coefficient of the transportation fails us, because $\mathbf{q}$ is the aggregation of the surface and ground flux:

$$\mathbf{q} = k(\mathbf{x}, t)\nabla h, \tag{7}$$

where $k(\mathbf{x}, t) = ah(\mathbf{x}, t)(a > 0)$ is the coefficient of transportation of water, ∇h is the variation of the level in x- and y-directions. This term allows us to take into account the sanding progress of sand in the lake or any geological component contributing to reduce the water level from the bottom of the basin. Finally, we obtain the following initial value problem for describing the lake level:

$$\frac{\partial h}{\partial t} - \nabla \cdot (k(h)\nabla h) + \alpha h^r = P \text{ in } \Omega \times]0, \infty), \tag{8a}$$

$$h = 0 \text{ in } \partial\Omega \times [0, \infty), \tag{8b}$$

$$h(\mathbf{x}, 0) = h_0(\mathbf{x}) > 0, \quad \forall \mathbf{x} \text{ in } \overline{\Omega} = \Omega \bigcup \partial\Omega. \tag{8c}$$

The system (8) is a special case of the problem model introduced by Diaz in [3]. For the sake of well-posedness and numerical approach of system (8), we admit the following regularity assumptions:

- $h \in C^1(0, \infty; L^2(\Omega))$;
- k, α and r are positive constants, the parameters of the model;

- $h_0 \in L^2(\Omega)$, and $h_0 = 0$ in $\partial\Omega$;
- $P \in C^1([0, \infty])$, $P \geq 0$.

2.2. Discretization of the lake surface

The spatial discretization of certain partial differential equations requires a discretization of the domain by cells where the unknown variable is located in the centers of cells.

Nowadays many pictures from lakes in general and Lake Chad in particular are available from satellites. We are concerned with the surface of water and it is known that the geometry of natural water areas is complex whence we need a preliminary treatment of the original picture in order to emphasize only two surfaces, the water and the land areas, and we include in land areas the swamps and forest.

The treatment of satellite image is done with Matlab image toolbox and some complementary programs. Here, we assume that the original surface of lake is domain Ω and the approximated domain is Ω_h such that the boundary of Ω_h is closed. Then, we imbed Ω_h in a rectangular domain $\widetilde{\Omega} \supset \Omega_h$. $\widetilde{\Omega}$ is discretized in rectangular finite elements (cells), thus we divide the cells in three subsets:

1. **F**, the subset of fluid cells which are entirely embedding in water area,
2. **B**, the subset of boundary cells which are both on land and water area, and
3. **L**, the subset of obstacle cells which are situated on land area.

It is shown in Figure 2 that when the number of finite element (cells) increases, then the image is neat. This phenomenon respects the main rule of pixel in the cameras.

A such decomposition of the domain required a special numerical scheme to approximate the solution of the system (8).

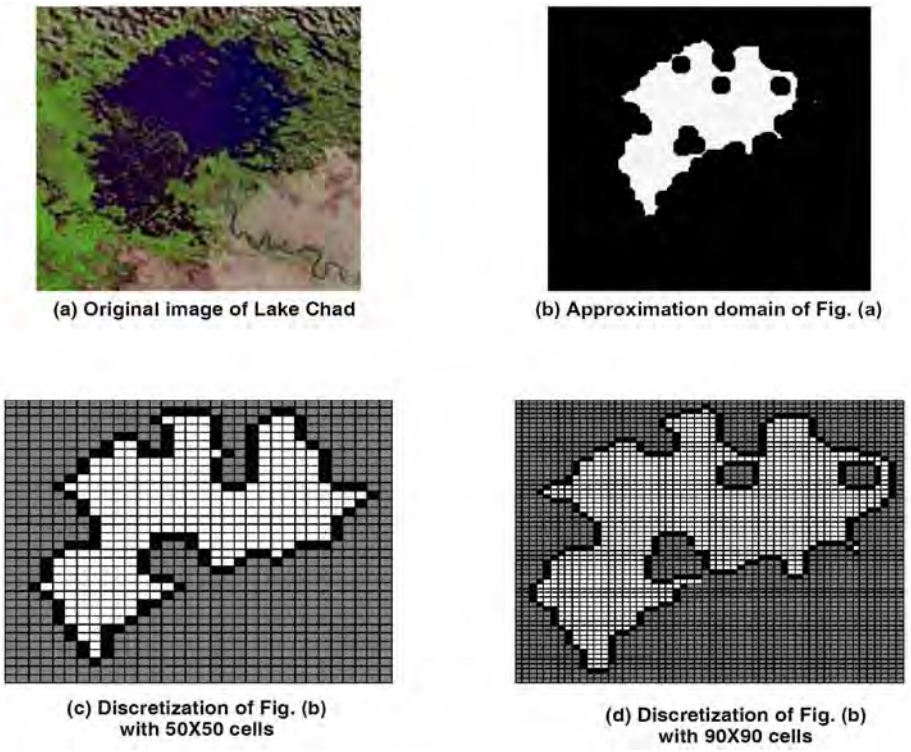


Figure 2. Process of treatment from original picture (Figure 2(a)) to discrete picture (Figure 2(c) and Figure 2(d)) via the approximation of domain $\widetilde{\Omega}$ (Figure 2(b)). Fluid Cells are in white color, boundary cells in black color and obstacle cells in grey color.

3. Numerical Scheme

We propose in this section to calculate an approximation to the solution of (8) with a numerical scheme suggested by Griebel et al. [6]. This scheme consists in using Donor-cell in space and Euler method in time.

3.1. Donor-cell scheme on space

We choose a rectangular grid with constant mesh sizes Δx and Δy in x- and y-directions, respectively. Let $i_{\max}$ and $j_{\max}$ be the total number of cells in x- and y-direction, respectively. We denote the grid counters as

$1 \le i \le i_{\max}$ and $1 \le j \le j_{\max}$ in x- and y-space, respectively. Equation (8a) can be rewritten as follows:

$$\frac{\partial h}{\partial t} - \frac{\partial}{\partial x}\left(k(h)\frac{\partial h}{\partial x}\right) - \frac{\partial}{\partial y}\left(k(h)\frac{\partial h}{\partial y}\right) + \alpha h^r = S. \tag{9}$$

The spatial derivatives

$$\frac{\partial}{\partial x}\left(k\frac{\partial h}{\partial x}\right) \text{ and } \frac{\partial}{\partial y}\left(k\frac{\partial h}{\partial y}\right)$$

are discretized at the cell center (i, j), $i = 1, ..., i_{\max}$, $j = 1, ..., j_{\max}$ by replacing the spatial derivative by the Donor-cell scheme. Details of the spatial discretization can be found in [6].

We shall use the notation $h_{i,j}$ for the numerical approximation to h at the center of the cell (i, j) and the subscripts $i \pm 1/2$, $j \pm 1/2$ denote the midpoints nodes in x-direction and y-direction, respectively. Therefore, the Donor-cell schemes for equation (8a) are the following:

$$\left[\frac{\partial}{\partial x}\left(k(h)\frac{\partial h}{\partial x}\right)\right]_{i,j} \simeq \frac{1}{\Delta x}\left(k_{i+1/2,j}\frac{h_{i+1,j} - h_{i,j}}{\Delta x} - k_{i-1/2,j}\frac{h_{i,j} - h_{i-1,j}}{\Delta x}\right) \tag{10}$$

and

$$\left[\frac{\partial}{\partial y}\left(k(h)\frac{\partial h}{\partial y}\right)\right]_{i,j} \simeq \frac{1}{\Delta y}\left(k_{i,j+1/2}\frac{h_{i,j+1} - h_{i,j}}{\Delta y} - k_{i,j-1/2}\frac{h_{i,j} - h_{i,j-1}}{\Delta y}\right). \tag{11}$$

3.2. Full discretization

We first introduce a maximum time T and an interval time $[0, T]$. Let Δt be a step size defined by $\Delta t = T/N$, where $N + 1$ is the total number of grid points. The notation h^n represents the approximation to $h(\mathbf{x}, t_n)$ for any $x \in \widetilde{\Omega}$ and $t_n = n\Delta t$, $0 \le n \le N$.

We set

$$h_{i,j}^{n+1} = h_{i,j}^{n} + \Delta t\left[\frac{1}{\Delta x}\left(k_{i+1/2,j}\frac{h_{i+1,j}^{n} - h_{i,j}^{n}}{\Delta x} - k_{i-1/2,j}\frac{h_{i,j}^{n} - h_{i-1,j}^{n}}{\Delta x}\right)\right.$$

$$+\frac{1}{\Delta y}\left(k_{i,j+1/2}\frac{h_{i,j+1}^{n} - h_{i,j}^{n}}{\Delta y} - k_{i,j-1/2}\frac{h_{i,j}^{n} - h_{i,j-1}^{n}}{\Delta y}\right)$$

$$\left. - \alpha h_{i,j}^{n} + P_{i,j}^{n}\right]; \quad \forall (i,\ j) \in \mathbf{F} \text{ and } 0 \le n \le N, \tag{12a}$$

$$h_{i,j}^{n} = 0; \quad \forall (i,\ j) \in \mathbf{B} \bigcup \mathbf{L} \text{ and } 0 \le n \le N, \tag{12b}$$

$$h_{i,j}^{0} = h_{i,j}^{0}; \quad \forall (i,\ j) \in \mathbf{F}. \tag{12c}$$

We will not go in depth in the stability condition of this scheme. It is well known that Donor-cell scheme is conditionally stable since the time step cannot be chosen independently of the spatial discretization. The stability criterion used is based on the Courant-Friedrichs-Lewy (CFL) conditions

$$\delta t < \frac{1}{2K_0\left(\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2}\right)}, \tag{13}$$

where K_0 and α are positive constants.

Another biologic criterion must be taken into account which is the positivity of the solution of (12). In the following section, we will illustrate three problems and point out their numerical solutions.

4. Numerical Experiments

In this section, we provide three computational examples. In the first one, we study the efficiency of (12) in term of relative error, CPU time and order of convergence in a regular domain. Note that, in this example, the maximum time is finite and the parameters k, α and r are known. The relative error is

calculated by means of the formula

$$E_r = \frac{\| H_h - H \|^2}{\| H \|^2} \tag{14}$$

and the order of convergence is computed from

$$s = \frac{\log\left(\frac{E_h}{E_{2h}}\right)}{\log(2)},$$

where H_h is the matrix solution obtained via (12) and H is the matrix solution of the exact solution. In the second example, we deal with a complex domain such as Figure 1, we present in color pictures the water level in the domain during the experience. Once again, the parameters k, α and r are known but not the exact solution. The aim of this example is to approximate the solution and compare to the exact one. The last example deals with the case of inverse problem. We consider that we have in our disposal data relative to the level of the water in a lake and we would like to identify the parameters k, α and r.

4.1. Example 1

Let $\Omega = [0, 1] \times [0, 1]$, the maximum time T is set to be 10 and we consider

$$k(x) = 1, \quad \alpha = 1, \quad r = 1.$$

The initial and the exact solutions are, respectively,

$$h_0(x, y) = xy(1 - x)(1 - y)$$

and

$$h(x, y, t) = xy(1 - x)(1 - y)e^{-2t}.$$

Numerical results are shown in Table 1. The first column is relative to the meshgrid, the second column is the relative error, the third is the CPU time and the last column of Table 1 is the order of convergence of the

method. The exact and the numerical solutions and the pointwise errors are shown in Figure 3.

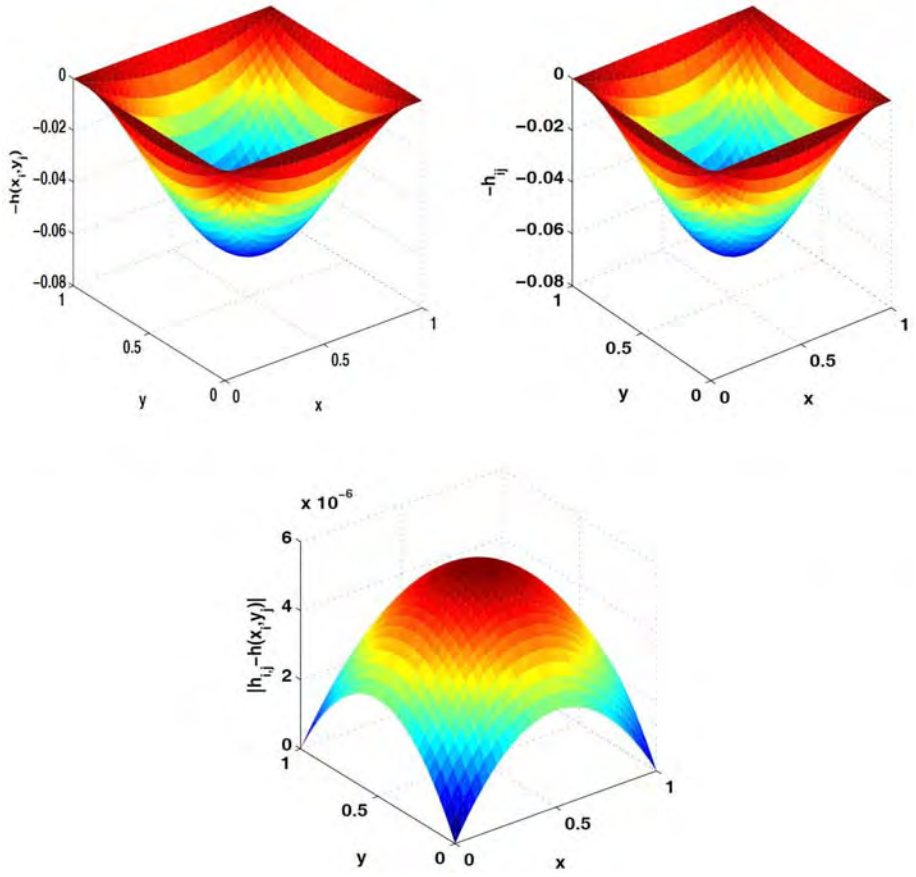


Figure 3. Example 1: The exact solution, the numerical solution and the pointwise error with $N = 50$.

Table 1. Example 1: $N \times N$ meshgrid, relative errors, CPU time (seconds) and order of convergence

N	E_r	CPU	s
5	$0.016 \cdot 10^{-2}$	0.023	-
10	$0.65336 \cdot 10^{-5}$	0.031	1.2921
15	$0.40944 \cdot 10^{-5}$	0.031	-
20	$0.2978 \cdot 10^{-5}$	0.031	1.332
25	$0.234 \cdot 10^{-5}$	0.031	-
30	$0.19267 \cdot 10^{-5}$	0.047	1.087
35	$0.16374 \cdot 10^{-5}$	0.063	-
40	$0.14235 \cdot 10^{-5}$	0.078	1.0652

4.2. Example 2

In this example, we consider the domain of lake like Figure 1 and the final time $T = 10$. We consider also the following parameters:

$$k(x) = 0.3, \quad \alpha = 2, \quad q = 0.8,$$

$$h_0(x, y) = xy(1 - x)(1 - y).$$

We present the water level graphs in the domain at time $t = 0; 2; 4; 6; 8; 10$ (Figure 4). These figures show the decreasing process of initial water level. This example confirms the drying process described by the model (8).

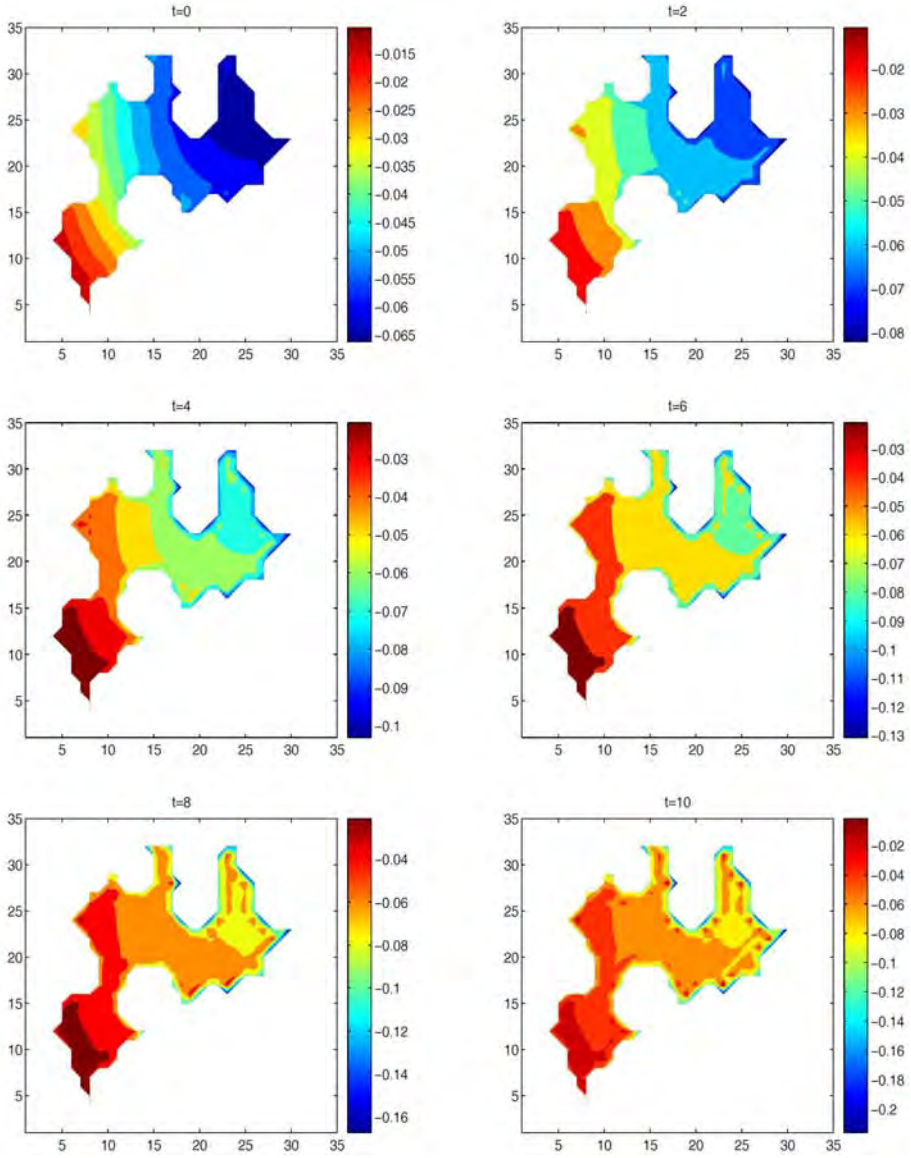


Figure 4. Example 2: The numerical solutions obtained by (12) on complex domain at time $t = 0; 2; 4; 6; 8; 10$.

4.3. Example 3

In this example, we are concerned by an inverse problem. We suppose that we have in our disposal data relative to the level of the lake, H_{obs}, collected in the time interval $[0, 10]$. The domain of the lake is the same as in Example 2. We would like to identify the parameters k, α and r which are constants. This is done by means of least square method [15] or sentinel method [16]. For this purpose, we introduce an operator of least squares method

$$J(k, \alpha, r) = \| H_{obs} - H(k, \alpha, r) \|^2 \tag{15}$$

such that

$$(\tilde{k}, \tilde{\alpha}, \tilde{r}) = \min_{(k, \alpha, r) \in R^3} J(k, \alpha, r), \tag{16}$$

where H is the approximated solution of (8) via Donor-cell algorithm (12). We consider H_{obs} the matrix solution of Example 1. We expect to recover the values of Example 2 parameters i.e.

$$k(x) = 0.3, \quad \alpha = 2, \quad q = 0.8.$$

The numerical results are shown in Table 2.

Table 2. Example 3: The approximated value of k, α and r are given according to the number of meshgrid and the CPU time

$\tilde{k}$	$\tilde{\alpha}$	$\tilde{r}$	N	CPU
0.3018	1.87	0.68	10	15.85
0.3005	1.7097	0.69	15	65.489
0.3003	1.7097	0.6898	20	68.952
0.3002	1.7635	0.6902	25	72.40
0.3001	1.7118	0.6900	30	77.361

5. Concluding Remarks

In this paper, we have presented a deterministic model to describe the drying-up process of a lake taking into account the complexity of the domain like Lake Chad surface. We have illustrated three numerical examples based on Donor-cells approximation to emphasize, firstly, the well known stability and convergence rate of the method, this is done in the first example. Secondly, in the second example we have shown the drying-up process of the model solution and finally in the last example we have identified the unknown parameters via some existing data. The paper points out numerical techniques to establish a deterministic model for any and in particular for Lake Chad if we have in our disposal the required data. This will be the next work of this investigation.

References

[1] T. Ayenew, Recent changes in the level of Lake Abiyata, central main Ethiopian Rift, Hydrological Sciences J. 47(3) (2002), 493-503.

[2] R W. Campbell, Lake Chad, West Africa: 1963, 1973, 1987, 1997, 2007 Earthshots: Satellite Images of Environmental Change, Reston, VA, US Geological Survey. Available at http://earthshots.usgs.gov (2008).

[3] J. I. Diaz, Qualitative study of nonlinear parabolic equations: an introduction, Extracta Mathematicae 16(3) (2001), 303-341.

[4] H. Gao, T. Bohn and D. P. Lettenmaier, Climate Change and Lake Chad: A 50-year Study from Land Surface Modeling, University of Washington, July 2010.

[5] H. Gao, T. J. Bohn, E. Podest, K. C. McDonald and D. P. Lettenmaier, On the causes of the shrinking of Lake Chad, Environmental Research Lett. 6(3) (2011), 034021.

[6] M. Griebel, T. Dornseifer and T. Neunhoeffer, Numerical Simulation in Fluid Dynamics. A Practical Guide, SIAM, Philadelphia, 1998.

[7] S. Howison, Practical Applied Mathematics: Modelling, Analysis, Approximation, Cambridge Texts in Applied Mathematics, 2005.

[8] S. Kebede, Y. Travi, T. Alemayehu and V. Marc, Water balance of Lake Tana and its sensitivity to fluctuations in rainfall, Blue Nile basin, Ethiopia, J. Hydrology 316(1-4) (2006), 233-247.

[9] J. Lemoalle, J.-C. Bader, M. Leblanc and A. Sedick, L'évolution récente du Lac Tchad: contexte général et données de base, Présenté en Conférence Invitée au Forum Mondial pour le Développement Durable "Pour la Sauvegarde du Lac Tchad", N'Djaména, 31 Octobre 2010.

[10] J. Lemoalle, J.-C. Bader and M. Leblanc, The variability of Lake Chad: hydrological modelling and ecosystem services, Proceedings of the 13th World Water Congress, Global Changes and Water Resources, 1-4 September 2008, Montpellier, France , pp. 1-15.

[11] P. G. Kumambala and A. Ervine, Water balance model of Lake Malawi and its sensitivity to climate change, Open Hydrology J. 4 (2010), 152-162.

[12] B. Pouyaud and J. Colombani, Les variations extrêmes au lac Tchad: l'assèchement est-il possible?, Annales de Géographie 98(545) (1989), 1-23.

[13] M. A. Roche, Évaluation des pertes du lac Tchad par abandon superficiel et infiltrations marginales, Cah. ORSTOM, Sér. Gèol. II(1) (1970), 67-80.

[14] E. Tate, J. Sutcliffe, D. Conway and F. Farquharson, Water balance of Lake Victoria: update to 2000 and climate change modelling to 2100, Hydrological Sciences J. 49(4) (2004), 574.

[15] A. Traoré, B. Mampassi and L. Longin, A least square spectral collocation formulation for solving PDEs on complex geometry domains, Inter. J. Appl. Math. Comput. 2(4) (2011), 9-22.

[16] A. Traoré, B. Mampassi and B. Saley, A numerical approach of the sentinel method for distributed parameter systems, Central European J. Math. 5(4) (2007), 751-763.

[17] G. Vuillaume, Bilan hydrologique mensuel et modélisation sommaire du regine hydrologique du lac Tchad, Cah. ORSTOM Hydrologie XVIII(1) (1981), 23-73.

[18] X. Yin and S. E. Nicholson, The water balance of Lake Victoria, Hydrological Sciences J. 43(5) (1998), 789-811.

ANNEXE 6

Article :

A. TRAORE, B. MAMPASSI, L. SOME, *A least-squares spectral collocation formulation for solving PDEs on complex geometry domains*, International Journal of Applied Mathematics and Computation Volume 2 (4), pp. 9--22, (2011)

International Journal of Applied Mathematics and Computation
Volume 2(4),pp 9–22, 2011
http://ijamc.psit.in

A least-squares spectral collocation formulation for solving PDEs on complex geometry domains

Aboubakari Traore[1], Benjamin Mampassi[1] and Longin Some[2]

[1] *Department of mathematics and computer sciences, Dakar University.*
Email: traboubakari@yahoo.fr and mampassi@hotmail.com

[2] *Department of mathematics and computer sciences, Ouagadougou University.*
Email: hsome@univouaga.bf

Abstract:
A least squares collocation scheme is used to solve PDEs defined on a complex geometry domain. Triangular finite elements are used to build a macro-mesh of the whole domain and Fekete points are used as collocation points. We combine together the standard least squares method and spectral method to compute the global solution. The assembling process of local solutions is explained through an algorithm. The success of this method is studied throughout a test problem. Numerical results prove the accuracy of the method which takes into account the boundary conditions.

Keywords: Least-squares formulation, Spectral elements, Fekete points, differentiation matrices.

1 Introduction

Computing solutions of partial differential equations (PDEs) on complex geometry domains is one of the important challenges in numerical analysis. Many classical numerical schemes suffer on the fact they do not take into account the complexity of the domain geometry.

It is well known that the finite element method (FEM) in its different variants is one of the most frequently used techniques to approximate the solutions of PDEs. The FEM is become more popular since the middle of the twentieth century mainly because of its applications by engineers to structural mechanics such that the analysis of pollutant transport in a fluid [3], [21], [30], the estimation of crustal deformation fields from GPS measurements [12], [15], [25], [26] and more generally some numerical solutions of Navier-Stokes equations [4], [13], [23]. One of the variant of the FEM is the Least-squares Finite Element Method (LSFEM) that gives accuracy solutions in the case of elliptic problems. But this method is not efficient for solving non linear and parabolic problems [5]. Alternatively to the LSFEM the Least Squares Spectral Collocation Method (LSSCM) was developed for solving some non linear PDEs. These methods have become an attractive class of methods for the numerical solutions of PDEs. LSSCM is particularly well suitable for solving problems that require coordinate transformations or tracking [22], [24], [29], [36]. However, it seems to require more computational effort and higher order continuity basis functions and more work in the matrix assembly. An other variant is the hp-Finite Element

Corresponding author: **Aboubakari Traore**

Method (hp-FEM). This version has been developed to take into account the boundary layers and the singularity of the solutions [16], [34], [37]. This method presents best abilities. For example the hp-FEM automatically refines the grid around the singularity and increases the polynomial order of approximation in regions where the solutions are smooth. The overall convergence of the method is very fast (exponential convergence).

Because the problem domain needs to be discretized into a mesh, these FEM versions suffer from drawbacks such as tedious meshing and re-meshing. The meshless method [2], [31],[32],[33], [42] has been proposed for the problem of computational mechanics in order to avoid the tedious meshing and re-meshing. The meshless method is used to establish a system of algebraic equations for the whole problem domain without the use of a predefined mesh.

Taking into account the problems met by the previous methods, FEMs has been developed in [40]. The domain is meshed by using a hybrid discretization consisting of triangular and quadrilateral sub-domains in order to provide great flexibility in refinement and unrefinement techniques. The spectral method is used over each finite element, in which the weak formulation of Galerkin is used to discretize the equations. An assembled matrix is easily defined but the difficulties for applying this method lies in the choice of suitable basis functions to interpolate the solutions.

In this paper we develop a scheme that combines LSCMs, [11], [22],[29] and Spectral collocation finite elements methods [14], [17],[18],[35]. Collocation points over an arbitrary triangle are generated using Fekete points [38].

We should also note that a lot of research for the LSSCM is done for Stokes and Navier-Stokes equations as well as for singular perturbation problems. The most important papers where LSSCM has been developed in different directions are given in [20], [27], [28]. For a general overview for spectral methods we should refer to [9]. Furthemore an adaptive LSSCM on triangular elements is presented in [19].

The paper is organized as follows. In Section 2 we describe the generating collocation points into a triangular element. In Section 3, we introduce the pseudospectral differentiation. The least squares collocation method is presented in section 4. The section 5 is devoted to numerical experiments and we conclude the paper with some remarks in section 6.

2 Triangular-collocation points

In this paper the whole domain is meshed by triangular elements. In this section we define collocation points in an arbitrary triangle by using Fekete points.

Let us define the standard triangle by

$$\widehat{T} = \{(r,s), -1 \leq r,s \leq 1;\ r+s \leq 0\} \tag{2.1}$$

and let us consider the Dubiner basis functions [14]

$$\phi_{ij}(r,s) = \left(\frac{1-s}{2}\right)^i \times p_i^{0,0}\left(\frac{2r+s+1}{1-s}\right) \times p_j^{2i+1,0}(s)$$

where the $p_j^{\alpha,\beta}(s)$ are the (α,β)-order Jacobi polynomials of degree j [1]. It is well known that the set of functions ϕ_{ij} , $0 \leq i,j \leq N$ and $i+j \leq N$ is an orthogonal basis of $P_N(\widehat{T})$, the space of polynomial of degree less than N. In all the following of this paper we shall write ϕ_k instead of ϕ_{ij}, $1 \leq k \leq (N+1)(N+2)/2$ for any arbitrary bijection $k \equiv k(i,j)$. Let us now consider the generalized Vandermonde matrix V whose components are $V_{ij} = \phi_j(z_i)$ for arbitrary points $z_k \in \widehat{T}$, $k = 1,...,\eta$, where we have set $\eta = (N+1)(N+2)/2$. Fekete points are the points $\widehat{z}_i$, $i = 1,...,\eta$ that maximize the determinant of V :

$$\max_{\{z_i\} \in \widehat{T}} |V(z_1, z_2, ..., z_N)| \tag{2.2}$$

In the framework of this paper Fekete points will be generate into an arbitrary triangular element using appropriate transformation map.

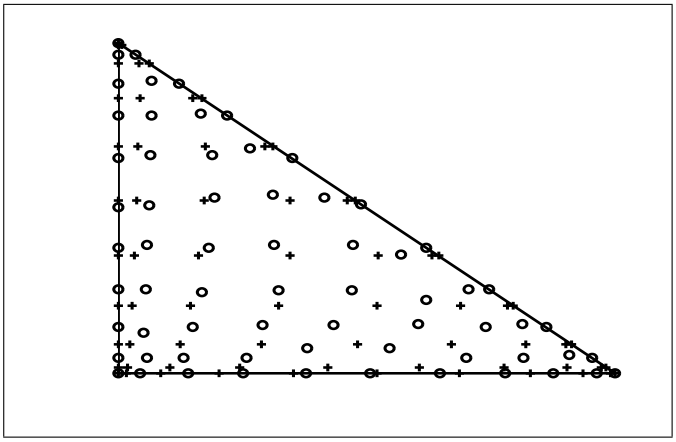


Figure 1: *The Lobatto triangle nodes (+) and associated Fekete nodes (o) over the reference triangle* $\widehat{T}$ *for* $N = 10$.

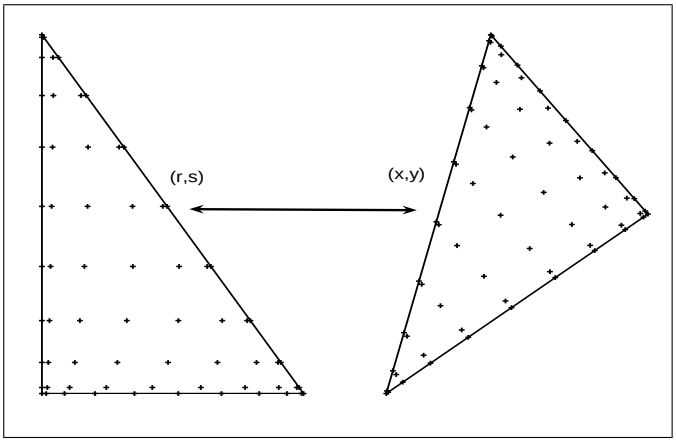


Figure 2: *Distribution of Fekete points from reference triangle (left) to arbitrary triangle (right) for* $N = 10$.

Fekete points are alternative collocation points to Gauss-Lobatto quadrature points. Clearly in the case of tensor-product domains such as lines or quadrangles, the Gauss- Lobatto quadrature points are well suitable for spectral approximation [7],[8]. However, in the case of the triangular domains, the Fekete points are commonly used in numerical methods to achieve both accurate high-order polynomial interpolation and quadrature properties [11],[38]. One can notice that the Fekete points are independent of the chosen basis but since we must compute numerically the inverse, it is important for the matrix to be well-conditioned.

3 Pseudo spectral Differentiation

Solving PDEs requires accurate approximation of derivatives. We want to provide a discrete differentiation matrix associated with Fekete collocation points on arbitrary triangles. We refer to [39]

in the case of one dimension and the standard quadrangle. In the case of triangular domains, singularities often arise near edges. Therefore both suitable collocation points and accurate numerical methods are required.

3.1 Differentiation over the reference triangle

For any continuous function $u(r,s)$ on the reference triangle $\widehat{T}$ we can write its spectral approximation in the space P_N by

$$u^N(r,s) = \sum_{k=1}^{\eta} U_k \ \phi_k(r,s) \tag{3.1}$$

where $\eta = (N+1)(N+2)/2$, and where the coefficients U_k are obtained by using collocation equations at Fekete points $\widehat{z}_m$

$$u(\widehat{z}_m) = \sum_{k=1}^{\eta} U_k \ \phi_k(\widehat{z}_m), \quad m = 1, 2, ..., \eta \tag{3.2}$$

Setting $U = (u(\widehat{z}_1),\ u(\widehat{z}_2),\ ...,\ u(\widehat{z}_\eta))^T$ and $C = (U_1,\ U_2,\ ...,\ U_\eta)^T$ the coefficients vector, we obtain

$$C = V^{-1} \times U \tag{3.3}$$

where V is the Vandermonde matrix for the Fekete points. According to (3.1), the derivatives in s and in r directions at Fekete collocation points $\widehat{z}_m$ are given by

$$\partial_r u^N(\widehat{z}_m) = \sum_{k=1}^{\eta} U_k \times \partial_r \phi_k(\widehat{z}_m)$$

and

$$\partial_s u^N(\widehat{z}_m) = \sum_{k=1}^{\eta} U_k \times \partial_s \phi_k(\widehat{z}_m)$$

respectively. We introduce two differentiation matrices V^r, V^s, of size $\eta \times \eta$ in $r-$direction respectively in $s-$direction whose components are $V^r_{ij} = \partial_r \phi_j(\widehat{z}_i)$ and $V^s_{ij} = \partial_s \phi_j(\widehat{z}_i)$. Denoting U_r respectively U_s the vector values of the differential approximation respectively in r and s directions at Fekete collocation points, we obtain

$$U_r = D^r \times U \qquad \text{and} \qquad U_s = D^s \times U \tag{3.4}$$

with

$$D^r = V^r \times V^{-1} \ \text{ and } \ D^s = V^s \times V^{-1} \tag{3.5}$$

We deduce the second order differentiation matrices over the reference triangle from (3.4) and (3.5)

$$D^{rr} = V^{rr} \times V^{-1}, \qquad D^{ss} = V^{ss} \times V^{-1}, \qquad D^{rs} = V^{rs} \times V^{-1} \tag{3.6}$$

where

$$V^{rr} = \partial_r V^r, \qquad V^{ss} = \partial_s V^s, \qquad V^{rs} = \partial_r V^s$$

3.2 Differentiation over an arbitrary triangle

Any derivative over an arbitrary triangle is derived from the reference triangle according to the bijective transformation such that:

$$u(r,s) = u(x_1(r,s), x_2(r,s)) \tag{3.7}$$

Applying the derivative rule in r (respectively s) direction, we have

$$\left\{ \begin{array}{rcl} \partial_r u & = & (\partial_r x_1)\,\partial_{x_1} u + (\partial_r x_2)\,\partial_{x_2} u \\ \partial_s u & = & (\partial_s x_1)\,\partial_{x_1} u + (\partial_s x_2)\,\partial_{x_2} u \end{array} \right. \tag{3.8}$$

Let us denote by U_{x_1} and U_{x_2} the vector values of the differential approximation respectively in x_1 and x_2 directions at Fekete collocation points. Then, from (3.5) and (3.8) we deduce

$$\begin{pmatrix} U_{x_1} \\ U_{x_2} \end{pmatrix} = G^{-1} \times \begin{pmatrix} D^r \\ D^s \end{pmatrix} \times U \tag{3.9}$$

where the matrix G is associated to the system (3.8) over Fekete collocation points. Setting $\mathbb{D} = G^{-1} \times \begin{pmatrix} D^r \\ D^s \end{pmatrix}$ then the differentiation matrices over an arbitrary triangle in $x-$direction, and $y-$direction are given by

$$D^x = \mathbb{D}(1:\eta,\ :\ ) \quad \text{and} \quad D^y = \mathbb{D}(\eta+1:2\eta,\ :\ ) \tag{3.10}$$

respectively. Second order differentiation matrices on an arbitrary triangle are derived from (3.6) and (3.8). Thus, there exists a matrix Q of size 3×3, obtained from the derivative of (3.8), such that

$$\left(\begin{array}{c} D^{xx} \\ D^{xy} \\ D^{yy} \end{array} \right) = Q \times \left(\begin{array}{c} D^{rr} \\ D^{rs} \\ D^{ss} \end{array} \right) \tag{3.11}$$

4 Least Squares spectral elements formulation

Collocation least squares methods (CLSM) also known as point least-squares or overdetermined collocation methods, have got a great success in elliptic equations solving [5],[6]. In this section we present them as an alternative method of standard Least Squares method applied to spectral collocation methods (LSSCM). The CLSM carried also a numerical contribution for solving some hyperbolic equations [10], [29], [41]. The parabolic problem case is not enough developed to our knowledge so we propose in this section the achievement on a test problem over a complex domain. The Fekete points have been used as collocation points over triangles.

4.1 The test problem

As a test problem, we consider the following parabolic equation

$$\left\{ \begin{array}{rcll} \Delta u + 2\pi^2 \times u & = & f & \text{in } \Omega \\ u_\Gamma & = & g & \text{on } \Gamma = \partial\Omega \end{array} \right. \tag{4.1}$$

where Ω is a plane domain that contains an obstacle taking into account the complexity of the domain geometry (see Fig. 3).

4.2 Triangulation and global assembling procedure

The differentiation procedure using finite element over a non standard domain requires the subdivision of the whole domain into finite elements (quadrilaterals, triangles) [40]. We have studied in the previous section the differentiation operators (the gradient and the Laplacian) over elementary triangles. We present now the procedure to assemble these elementary operators into a global operator by means of an operator that assembles the local coefficients into the global coefficients and ensure C^0 continuity.

Our numerical scheme is based on computing a local solution u_{loc} over each elementary triangles and then the global solution is obtained using an assembly procedure. To illustrate this global

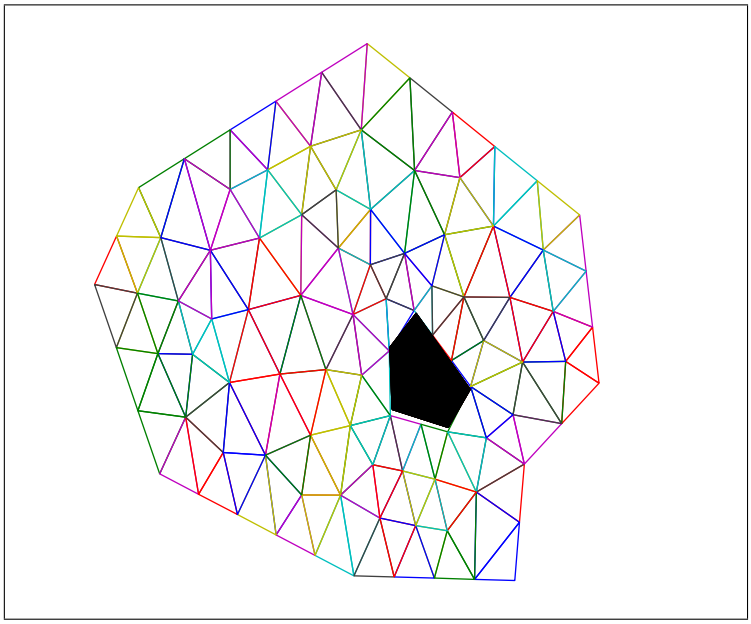


Figure 3: *Triangular macromesh of computational domain.*

assembly procedure, we consider a global domain made up of two triangles as shown in the figure below.

The figure (4) illustrates the local and global numbering of a domain containing two triangular elements. For example we have taken $N = 3$ for the expansion order. This technique will allows us to avoid the repeated collocations points in global assembly. In the examples above, the total number of freedom in local numbering is $N_{Loc} = 12$ and the global numbering is $N_{glob} = 9$. Let u be a continuous function defined over such a domain and $\widehat{u}_g$ the global vector value of u calculated over the global collocation points

$$\widehat{u}_g = \left(\widehat{u}_g^1, \widehat{u}_g^2, ..., \widehat{u}_g^9\right)^T \tag{4.2}$$

and $\widehat{u}_1$ and $\widehat{u}_2$ respectively the vector value of u over the first (respectively the second) triangular element

$$\widehat{u}_1 = \left(\widehat{u}_1^1, \widehat{u}_1^2, ..., \widehat{u}_1^6\right)^T \text{ and } \widehat{u}_2 = \left(\widehat{u}_2^1, \widehat{u}_2^2, ..., \widehat{u}_2^6\right)^T \tag{4.3}$$

In this case the global assembled matrix A is a (12×9) matrix such that

$$\begin{bmatrix} \widehat{u}_1 \\ \widehat{u}_2 \end{bmatrix} = A \times \widehat{u}_g \tag{4.4}$$

The mesh of the domain above can be done by the program "dismesh_2d" available in Matlab software, this returns the matrix of p coordinates of each edge and the matrix of numbering of triangulation. Let us consider an elementary triangle k in Ω, the problem (4.1) is linear so there exists a matrix H^k such that

$$H^k \times U^k + 2\pi^2 U^k = f^k \tag{4.5}$$

for $1 \leq k \leq n_t$, where n_t is the total number of triangles in Ω. We deduce the system

$$\begin{pmatrix} H^1 & \ddots & 0 \\ \ddots & \ddots & \ddots \\ 0 & \ddots & H^{n_t} \end{pmatrix} \times \begin{pmatrix} U^1 \\ \vdots \\ U^{n_t} \end{pmatrix} + 2\pi^2 \times \begin{pmatrix} U^1 \\ \vdots \\ U^{n_t} \end{pmatrix} = \begin{pmatrix} f^1 \\ \vdots \\ f^{n_t} \end{pmatrix} \tag{4.6}$$

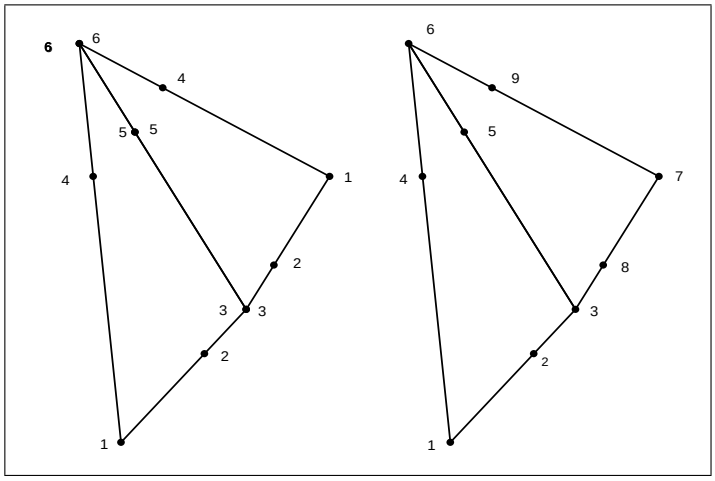


Figure 4: *Local numbering (left) and global numbering (right).*

Let us denote by

$$\mathbb{H} = \begin{pmatrix} H^1 & \ddots & 0 \\ \ddots & \ddots & \ddots \\ 0 & \ddots & H^{n_t} \end{pmatrix} \tag{4.7}$$

and

$$U_{Loc} = \left(U^1,\ ...,U^{n_t}\right)^T \tag{4.8}$$

$$F_{Loc} = \left(f^1,\ ...,\ f^{n_t}\right)^T \tag{4.9}$$

According to the relation (4.4), there exists an assembled matrix Z such that:

$$U_{Loc} = Z \times U_{glob} \tag{4.10}$$

Then, the relation (4.6) becomes

$$\mathbb{H} \times Z \times U_{glob} + 2\pi^2 \times Z \times U_{glob} = F_{Loc} \tag{4.11}$$

Hence, for simplicity, we deduce from (4.11) a differentiation matrix $\mathbb{D}$ such that:

$$\mathbb{D}{\times}U_{glob} + 2\pi^2 \times U_{glob} = F_{glob}$$

where we have set

$$\mathbb{D} = inv(Z^T \times Z) \times Z^T \times \mathbb{H}{\times}Z \tag{4.12}$$

It is important to note that the definition of $\mathbb{D}$ is justified by the fact that the matrix Z is not square. For boundary conditions, there exists a matrix $\mathbb{B}$ such that

$$\mathbb{B}{\times}U_{glob} = G_{glob} \tag{4.13}$$

The operator of least squares method is then derived from (4.11 and 4.13)

$$\mathcal{L}(U) = \widehat{\alpha}\left\|\mathbb{D}{\times}U + 2\pi^2 \times U - F_{glob}\right\|^2_{\mathbb{R}^{n_p}} + \widehat{\beta}\left\|\mathbb{B} \times U - G_{glob}\right\|^2_{\mathbb{R}^{n_b}} \tag{4.14}$$

where the weights $\widehat{\beta}$ and $\widehat{\alpha}$ can be used to adjust the relative importance of the terms in the functional. Here $\|.\|_{\mathbb{R}^{n_p}}$ and $\|.\|_{\mathbb{R}^{n_b}}$ denote l^2-norm respectively over $\mathbb{R}^{n_p}$ and $\mathbb{R}^{n_b}$. The integers

n_p and n_b represent the number of collocation points on the whole domain and the number of boundary collocation points. Finally, one obtains the numerical solution of the problem (4.1) by solving the following problem

$$\min_{U \in \mathbb{R}^{n_p}} \mathcal{L}(U) \tag{4.15}$$

We summarize the algorithmic scheme in (Fig.5).

(1) Read the parameters:

- n : Mesh parameter.
- N : degree of interpolation polynomials.

(2) Compute the vectors and matrices associated to the meshing:

- $\mathbb{B}$: Matrix which selects the boundary points, computed according to (eq.26).
- $\mathbb{D}$: Discret differential operator, computed as (eq.25).
- x_{loc} : Vector of local collocation points, obtained according to (eq.2).
- $x_{glob} = Z \times x_{loc}$: Vector of global collocation points
- $x_{bord} = D_{bord} \times x_{glob}$: Vector of boundary points
- $\widehat{g}$: Boundary values computed using (eq.29) evaluated at each point of x_{bord}.

(3) Define the coefficients $\widehat{\beta}$ and $\widehat{\alpha}$,

(4) Define the least squares operator (eq.27):

$$\mathcal{L}(U;g) = \widehat{\alpha} \left\| \mathbb{D} \times U + 2\pi^2 \times U \right\|^2_{R^{N_p}} + \widehat{\beta} \left\| \mathbb{B} \times U - \widehat{g} \right\|^2_{R^{N_b}} \tag{26}$$

(5) Choose an arbitrary value U_0 for iterative process

(6) Compute by an iterative method the R^{N_p} minimizing problem

$$\mathcal{L}\left(U^*;g\right) = \min_{U \in R^{N_p}} \mathcal{L}\left(U;g\right)$$

Figure 5: *Algorithm for the Least Squares Spectral Elements Method.*

5 Computational Experiments

For our numerical experiment we consider the test problem (4.1) from which the exact solution is

$$u_e = \sin(\pi x)\sin(\pi y) \tag{5.1}$$

when we have taken $f = 0$ and the data of the function g called G_{glob} in (4.14) are available by computing the value of (5.1) at boundaries points. All calculations are performed with a fixed mesh size by varying the degree N of the interpolation polynomial. The aim of this experience is to analyze the efficiency of the algorithmic scheme (Fig.5).

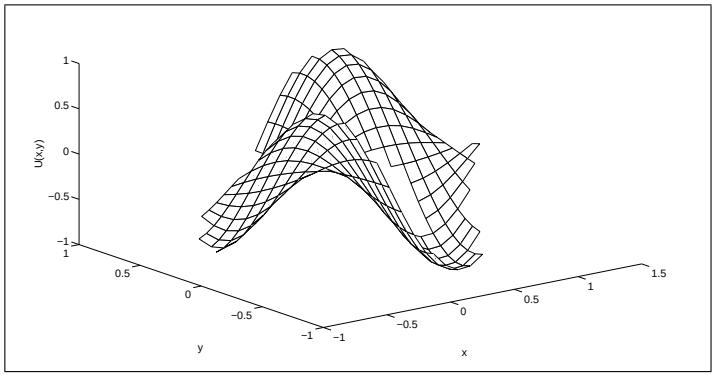


Figure 6: *The exact solution (5.1)*

Figures 7, 8 and 9 illustrate the solution of problem (4.1) for $N = 2, 4$ and 6 respectively. These graphics show the convergence process of approached solutions toward the exact solution and take better account of boundary conditions. The gradient field in figures (10-12) shows the convergence of the solution in regions with high concentration gradient. In figures (13-15), we evaluate the absolute error committed when approximating first and second order partial derivatives. Thus, it is clear that these errors become as small as the number N of collocation points increases. Starting from $N = 4$, the approached solution is very close to the exact one.

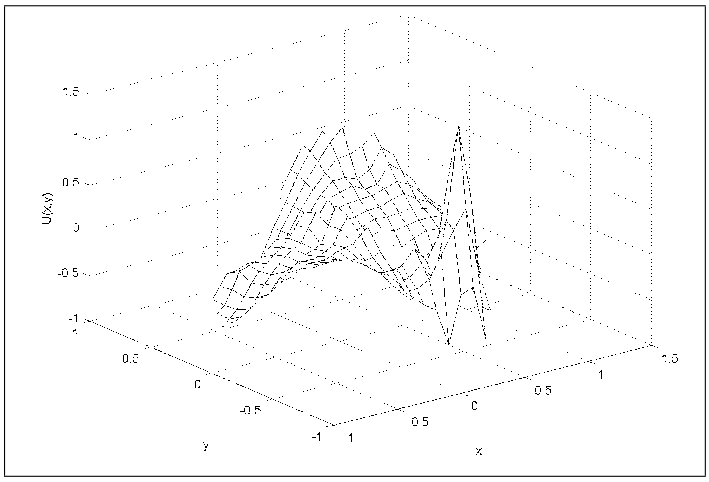


Figure 7: *Numerical solution for* $N = 2$*, and* $\hat{\alpha} = \hat{\beta} = 1$.

6 Concluding remarks

A least squares collocation scheme for solving PDEs over a complex domain is presented. Using triangular finite elements and Fekete points, the assembling process of global solution has been quite easy. The macro-mesh of complex domain by triangles has enabled to take into account the boundary conditions. Numerical simulations on a test problem have confirmed the high accuracy of our spectral least-squares scheme.

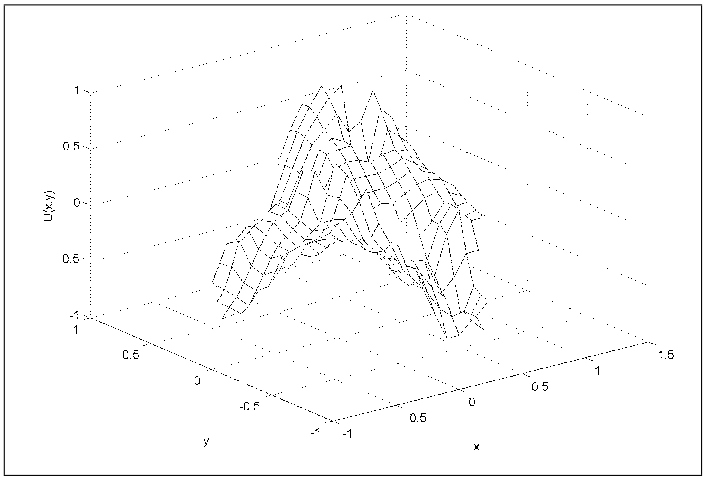


Figure 8: *Numerical solution for* $N = 4$*, and* $\hat{\alpha} = \hat{\beta} = 1$.

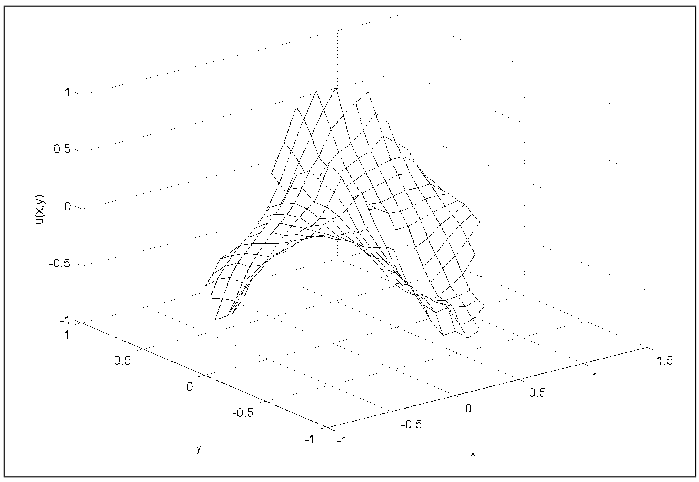


Figure 9: *Numerical solution of (4.1) with* $N = 6$*, and* $\hat{\alpha} = \hat{\beta} = 1$.

References

[1] Abramowitz M., Stegun I.A.(ed.), Handbook of mathematical functions with formulas, graphs, and mathematical tables, Wiley-Interscience (1972).

[2] Atluri S. N., and Shen S. P., The Meshless Local PPetrov–Galerkin (MLPG) Method, Tech Science Press, Encino, CA, pp. 93–214, (2002).

[3] Baptista A.E., Adams E.E., and Stolzenbach K. D., Eulerian-Lagrangian Analysis of Pollutant Transport in Shallow Water, Ralph M. Parsons Laboratory, MIT, Rpt. No. 296, (1984).

[4] Bentley L. R., and Pinder G. F., Solution of the Advective-Dispersive Transport Equation using a Least Squares Collocation, Eulerian-Lagrangian Method, John Wiley & Sons, Inc. Numerical Methods for Partial Differential Equations, 5, pp. 227-240 (1989).

[5] Bochev P. and Gunzburger M., least squares finite element methods, Prodeeding of the International Congress of Mathematicians, vol III, pp.1137-1162, (2006).

[6] Bochev P. and Gunzburger M., least-squares finite element methods for first order elliptic system, international journal of numerical analysis and modeling. volume 1, number 1, pp. 49-64.

[7] Bos L., On certain configurations of points in Rn which are uniresolvant for polynomial interpolation, J. Approx. Theory, 64 , pp. 271-280, (1991).

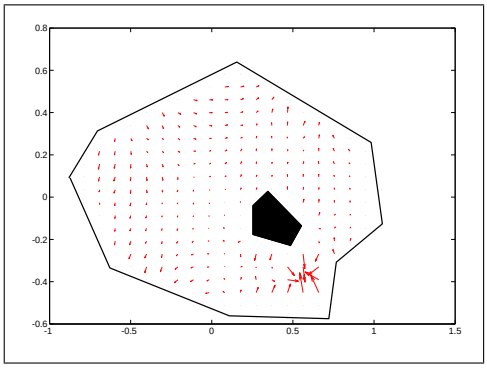


Figure 10: *Gradient of solution through the domain for* $N = 2$, *and* $\hat{\alpha} = \hat{\beta} = 1$.

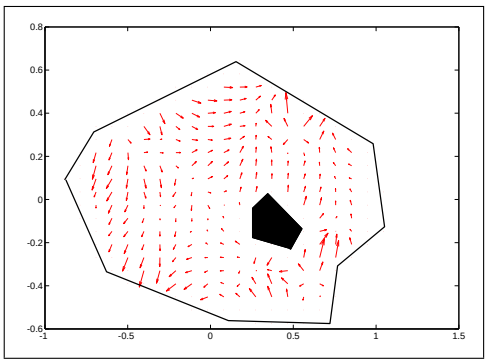


Figure 11: *Gradient of solution through the domain for* $N = 4$, *and* $\hat{\alpha} = \hat{\beta} = 1$.

[8] Bos L., Taylor M.A., Wingate B.A., Tensor product Gauss-Lobatto points are Fekete points for the cube, Math. Comp., 70 , pp. 1543-1547,(2001).

[9] Canuto C., Hussaini M.Y., Quarteroni A., and Zang T.A., Spectral methods: Evolution to complex geometries & applications to fluid dynamics (Scientific computation), Springer, Scientific Computation, (2007)

[10] Chang C. L., Gunzburger M., A subdomain Galerkin/ least square method for first order elliptic systems in the plane, SIAM J. Numer. Anal . 27, pp.1197.1211, (1990).

[11] Chen Q. and Babuska I., Approximate optimal points for polynomial interpolation of real functions in an interval and in a triangle, Comput. Methods Appl. Mech. Engrg., 128 , pp. 405–417, (1995).

[12] Cocard M., Kahle H.G., Peer Y., Geiger A., Veis G., Felekis S., Paradissis D., and Billiris H.. New constraints on the rapid crustal motion of the Aegean region: recent results inferred from GPS measurements (1993–1998) across the West Hellenic Arc, Greece, Earth planet. Sci. Lett., 172, pp. 39–47 (1999).

[13] Douglas J. and Russell T. F., Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures, SIAM J . Numer. Anal.. 19, 871-885(1982).

[14] Dubiner M., Spectral methods on triangles and other domains, J. Sci. Comput., 6 , pp.345–390, (1993).

[15] Egli R., Geiger A., Wiget A. and Kahle1 H.-G., A modified least-squares collocation method for the determination of crustal deformation: first results in the Swiss Alps. Geophys. J. int 168, pp.1-12, (2007).

[16] Garcia-Castillo L. E., Pardo D., Demkowicz L. F., A two-dimensional sel-adaptive hp-adaptive finite element method for the characterization of waveguide discontinuities. Part I: Waveguide theory and finite element formulation, Computer Methods in Applied Mechanics and Engineering, Volume 196, Issues 49-52, pp. 4823-4852 (2007).

[17] Heinrichs W., An Adaptive Spectral Least-Squares Scheme for the Burgers Equation, Springer Netherlands ,Volume 44, Number 1, pp. 1017-1398 , (Jan 2007)

[18] Heinrichs W., Least-Squares Spectral Collocation with the Overlapping Schwarz Method for the Incompressible Navier–Stokes Equations, Springer, Numerical Algorithms, Volume 43, Number 1, pp. 61-73(13), (September 2006).

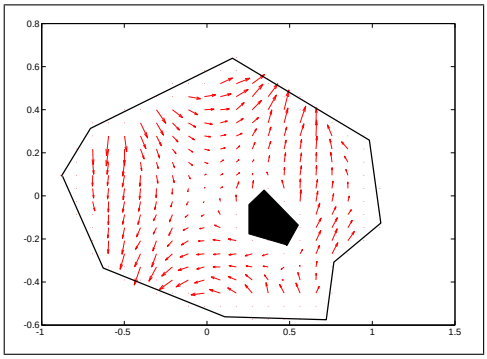


Figure 12: *Gradient of solution through the domain for* $N = 6$, $\hat{\alpha} = \hat{\beta} = 1$.

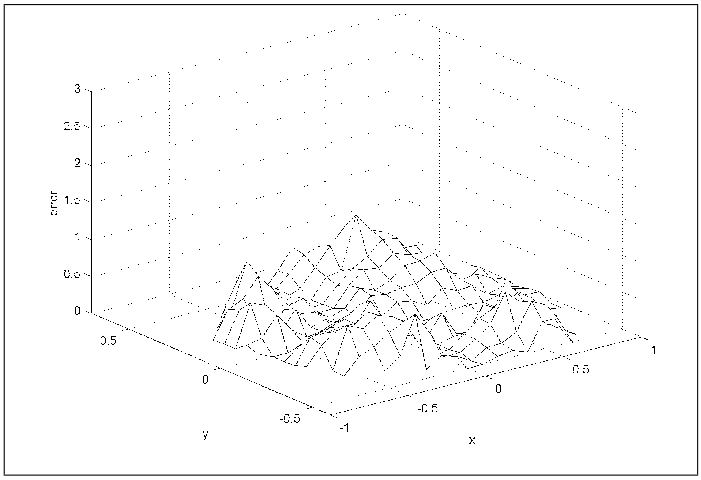


Figure 13: *Distribution of absolute error for* $N = 2$, *and* $\hat{\alpha} = \hat{\beta} = 1$.

[19] Heinrichs W., An adaptive least-squares spectral collocation method with triangular elements for the incompressible Navier-Stokes equations , Journal of Engineering Mathematics J. Eng. Math. 56(3), pp. 337-350 (2006)

[20] Heinrichs W. Least-squares spectral collocation for discontinuous and singular perturbation problems , Journal of Computational and Applied Mathematics, 157(2), pp. 329-345 (2003)

[21] Holly F. M.,and Komatsu T., Derivative approximations in the two-point fourth order method for pollutant transport, in Proceedings of the Conference on Frontiers in Hydraulic Engineering, ASCE, MIT, Cambridge, , pp. 349-355, (1983).

[22] Houstis E. N., and Lynch R. E., Evaluation of numerical methods for elliptic partial differential equations, J . Comp. Phys., 27, pp.323-350 (1978).

[23] Huffenus J. P. and Khaletzky D., The Lagrangian approach of advective term treatment and its application to the solution of Navier-Stokes equations, Int. J. Numer. Meths. Fluids, 1, (1981),pp.365-387.

[24] Joos B., The Least Squares Collocation Method for Solving Partial Differential Equations, Ph.D. dissertation, Princeton University, (1986).

[25] Kahle H.-G., Müller M.V., Geiger A., Danuser G., M¨uller S., Veis G., Billiris H. & Paradissis, D.. The strain field in northwestern Greece and the Ionian Islands: results inferred from GPS measurements, Tectonophysics, 249, pp.41–52, (1995).

[26] Kahle H.G., Cocard M., Peter Y., Geiger A., Reilinger R., Barka A., & Veis G., GPS-derived strain rate field within the boundary zones of the Eurasian, African, and Arabian plates, J. geophys. Res., 105, pp. 23 353–23 370, (2000).

[27] Kattelans T., and Heinrichs W., Conservation of mass and momentum of the least-squares spectral collocation scheme for the Stokes problem Journal of Computational Physics, 228(13), pp. 4649-4664 (2009)

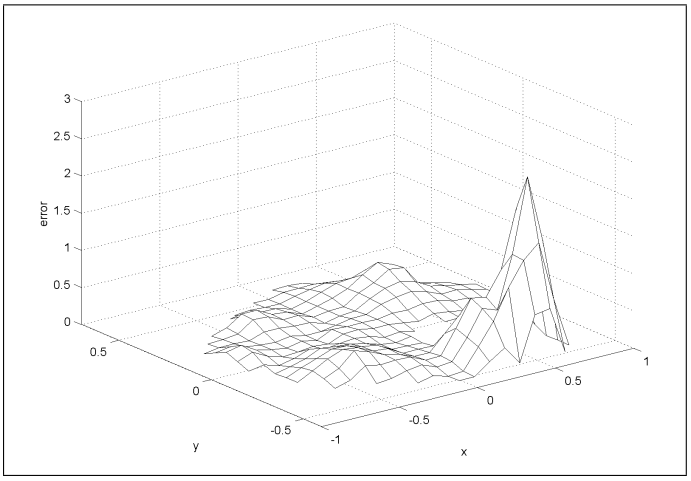


Figure 14: *Distribution of absolute error for* $N = 4$*, and* $\hat{\alpha} = \hat{\beta} = 1$.

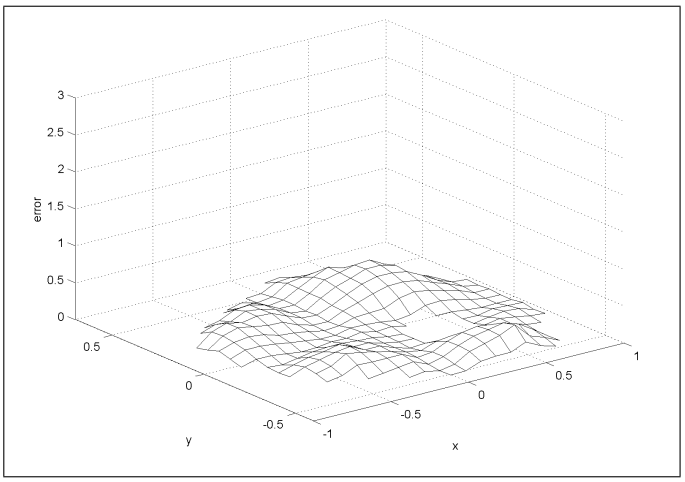


Figure 15: *Distribution of absolute error for* $N = 6$*, and* $\hat{\alpha} = \hat{\beta} = 1$.

[28] Kattelans T., and Heinrichs W., A direct solver for the least-squares spectral collocation system on rectangular elements for the incompressible Navier-Stokes equations , Journal of Computational Physics, 227(9), pp. 4776-4796 (2008)

[29] Laible J. and Pinder G., Least squares collocation solution of differential equation on irregularly shaped domains using orthogonal meshes, Numer. Meth PDE's, 5 pp. 347-361, (1989).

[30] Lapidus L. and Pinder G., Numerical Solution of Partial Differential Equations in Science and Engineering, John Wiley & Sons, (1999).

[31] Liu, G. R., Mesh Free Methods, CRC Press, Boca Raton, FL, pp. 67–248, (2003).

[32] Liu, G. R., and Gu, Y. T., An Introduction to Meshfree Methods and their Programming, Springer, New York, pp.54–144, (2005).

[33] Liu, L. H., Meshless Local PPetrov–Galerkin Method for Solving Radiative Transfer Equation, Journal of Thermophysics and Heat Transfer, Vol. 20, No. 1, pp. 150–154, (2006)

[34] Pardo D., Garcia-Castillo L. E., Demkowicz L. F., Torres-Verdin C., A Two-Dimensional Self-Adaptive hp Finite Element Method for the Characterization of Waveguide Discontinuities. Part II: Goal-Oriented hp-Adaptivity, Elsevier, Amsterdam, PAYS-BAS, vol. 196, no 49-52, pp. 4811-4822 (2007).

[35] Pasquetti R., Rapetti F., Spectral element methods on triangles and quadrilaterals: comparisons and applications, Volume 198, Issue 1, , Pages 349-362, (2004).

[36] Rosenblueth E., Physical control of numerical solution of parabolic equations, Eng. Anal., 2, pp.107-110 (1985).

[37] De Sterck H., Manteuffel T. A., McCormick S.F., Nolting J., Ruge J., and Tang L., Efficiency-based h- and hp-refinement strategies for finite element methods, John Wiley & Sons, Ltd. Numerical Linear Algebra with Applications Numer. Linear Algebra Appl. ; 00: pp.1–25 (2007).

[38] Taylors M. A., Wingate B. A., and Vincent R. E., An algorithm for computing FEKETE points in the triangle, SIAM J. NUMER. ANAL. Society for Industrial and Applied Mathematics, Vol. 38, No. 5, pp. 1707–1720, (2000).

[39] Trefethen L.N., Spectral Method in Matlab, Siam (2000).

[40] Warburton T.C., Sherwin S.J., and Karniadakis G.E., Basis Functions for Triangular and Quadrilateral High-Order Elements, SIAM Journal on Scientific Computing Volume 20 , Issue 5 pp. 1671 - 1695, (September 1999)

[41] Zeitoun D. and Pinder G., A least squares approach for solving remediation problems of contaminated aquifers, Numerical methods in water resources, 4 pp. 329-335, (1989).

[42] Zhang X., and Liu Y., , Meshless Methods, Tsinghua University Press, Beijing, pp. 144–176, (2004).