

Publication pour la présente candidature LAFPT

Titre	Analyse spatio-temporelle de la Fièvre de la Vallée du Rift au Ferlo (Sénégal)
Auteurs	Fanta Bouba, Alassane Bah, Christophe Cambier, <u>Samba Ndiaye</u> , Jacques-André Ndione; Maguelonne Teisseire.
Année	2014
Référence	12ème colloque africain sur la recherche en informatique et mathématiques appliquées (CARI 2014)
Editeur	CARI
Pages	317 – 324
DOI	hal-01062320
URL	http://www.cari-info.org/cari_2014/p33.pdf
Index	
ISBN	
Encadreur	Oui
Extrait d'une thèse	Oui



Accéder au rapport CARI-2014 / Access to CARI'2014 report

PROCEEDINGS

N	Title of the paper and authors	Page
31	Towards a Multi-Model Software Platform for Crops Growth Simulation Dieu Donne Okalas Ossami, Latyr Senghor	293 - 303
32	Forward kinematics modeling of a Compact Bionic Handling Assistant Manipulator Achille Melingui, Rochdi Merzouki, Jean Bosco Mbede	305 - 316
33	Analyse spatio-temporelle de la Fièvre de la Vallée du Rift au Ferlo (Sénégal) Fanta Bouba, Alassane Bah, Samba Ndiaye, Jacques André Ndione, Christophe Cambier, Maguelonne Teisseire	317 - 324
34	Un modèle de documents stable par projections pour l'édition coopérative asynchrone Maurice Tchoupé Tchendji, Marcellin Atemkeng, Rodrigue Djeumen	325 - 332
35	Evaluation des requêtes avec préférences structurelles sur les documents XML Maurice Tchoupé Tchendji, Lionnel Tadonfouet	333 - 341
36	Vers une approche de veille épidémiologique adaptée aux pays en voie de développement : Cas de la schistosomiase au Sénégal Gaoussou Camara, Sylvie Despres, Moussa Lo	343 - 350
37	Architecture d'un système de gestion des connaissances de la médecine traditionnelle : ontoMEDTRAD Kouame Appoh, INPHB, Moussa Lo, Marcellin Brou Konan	351 - 362

Activer Windows

Analyse spatio-temporelle de la Fièvre de la Vallée du Rift au Ferlo (Sénégal)

Fanta Bouba¹ et Alassane Bah¹ et Samba Ndiaye¹ et Jacques-André Ndione² et
Christophe Cambier³ et Maguelonne Teisseire⁴

¹ UMI 209, UMMISCO-UCAD et UCAD, Dakar, Sénégal

² CSE et LPA/UCAD, Sénégal

³ UMI 209, UMMISCO-IRD et Université Pierre et Marie Curie, France

⁴ UMR TETIS et Université de Montpellier II, France

bouhafanta@gmail.com ; alassane.bah@gmail.com ; samba.ndiaye@ucad.edu.sn ;
jacques-andre.ndione@cse.sn ; christophe.cambier@upmc.fr ;
maguelonne.teisseire@teledetection.fr

RÉSUMÉ. La Fièvre de la Vallée du Rift (FVR) est une anthro-poo-zoonose répandue dans les régions tropicales. Au Sénégal, elle est particulièrement ressentie dans la zone du Ferlo où il est noté une forte présence de mares partagées par les hommes, le bétail et les vecteurs. Dans le cadre des études menées sur les facteurs environnementaux qui favorisent son déclenchement et sa propagation, nous travaillons sur le processus décisionnel pour évaluer les impacts, les interactions et faciliter le suivi spatio-temporel de la FVR. Cet article propose les résultats des analyses faites sur les caractéristiques des mares en se basant sur les motifs spatio-temporels. Cette démarche présente un intérêt pour la mise en évidence des relations entre les données environnementales et les vecteurs de transmission de FVR.

ABSTRACT. The Rift Valley Fever (RVF) is an anthro-po-zoonosis spread in tropical areas. In Senegal, it is particularly felt in the Ferlo area where a strong presence of ponds shared by humans, cattle and vectors is noted. As part of the studies carried out on the environment factors which favour its start and propagation, we work on the decision making process to evaluate the impacts, the interactions and to facilitate spatio-temporal monitoring of RVF. The present paper proposes the results of the analyses made on the characteristics of the ponds based on spatio-temporal patterns. This approach presents some advantage in revealing the relationship between environmental data and RVF transmission vectors.

MOTS-CLÉS: fouille de données, système décisionnel, modèle multidimensionnel, Fièvre de la Vallée du Rift, motifs spatio-temporels

KEYWORDS: datamining, decision-making system, multidimensionnall modelling, Rift Valley Fever, spatio-temporal patterns

1. Introduction

La Fièvre de la Vallée du Rift (FVR) est une arbovirose qui se transmet par des vecteurs. Malgré un taux de mortalité peu élevé, cette maladie, vue comme une « arme biologique et économique », a « un grand pouvoir de diffusion et une gravité particulière » [4]. En effet, la dissémination de son virus peut se faire par aérosol et les oeufs des vecteurs (*Aedes vexans*, *Culex poicilipes*) sont résistants à la durée et à la sécheresse [8]. C'est pourquoi elle figure sur la liste commune des maladies à plusieurs espèces de l'Organisation Internationale pour la Santé Animale [26]. La compréhension du processus d'émergence et de propagation, telle qu'abordée actuellement, n'offre qu'une vue cloisonnée parcellaire. Pourtant, d'importants travaux [20,17], ont permis de montrer l'impact des facteurs tels que le climat, la transhumance des troupeaux, l'hydrologie sur l'apparition et la persistance de la FVR. Cette complexité d'analyse nécessite l'implication de tous les acteurs concernés dans une dynamique pluridisciplinaire. Cette approche, basée sur l'analyse multicritères et multidimensionnelle, permet de proposer des mesures de gestion du risque sanitaire en maîtrisant les impacts des facteurs environnementaux et leurs interactions. Dès lors, nous proposons d'adopter une approche décisionnelle afin d'offrir aux experts métiers et décideurs des outils d'analyse, rapportée aux dynamiques des populations animales et humaines. Aussi, pour la découverte d'informations, nouvelles et pertinentes, et afin de comprendre les interactions entre les différents objets et faire des prévisions, il est nécessaire de recourir à des techniques de fouille de données dans un objectif de classification et de prédiction. Ces techniques sont appliquées afin d'évaluer la pertinence des paramètres retenus pour le suivi spatio-temporel de la FVR. Dans la suite de cet article, la section 2 est consacrée à une description plus détaillée de la maladie étudiée et de la fouille de données appliquée en épidémiologie. Nous décrivons, en section 3, les données utilisées pour cette expérimentation, leur structure dans l'entrepôt de données et les motifs spatiaux temporels et les données d'expérimentation. Dans la section 4, nous présentons des extraits des analyses corrélées de caractérisation des mares temporaires de la zone d'étude. La section 5 conclut cette démarche en discutant les choix réalisés et en proposant des pistes d'analyse pour la suite de nos travaux.

2. Maladie de la FVR

La Fièvre de la vallée du Rift (FVR) est une maladie infectieuse et virale, qui affecte les humains et les animaux [19]. Le nom de la maladie est originaire de la zone dans laquelle le virus a été isolé pour la première fois ; il s'agit de la vallée du Rift au Kenya. Cette maladie, véhiculée par des arthropodes vecteurs [6], est causée par un virus du genre Phlebovirus de la famille des Bunyaviridae. Au Sénégal, cette maladie occupe une place importante dans le Ferlo. Selon [5], la propagation de la FVR, identifiée dans la zone du Ferlo (figure 1), est liée au cycle de vie des moustiques et au cycle

d'évolution des mares.

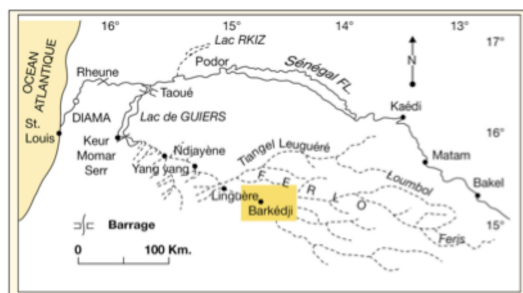


Figure 1. Localisation de la zone d'étude, adaptée de [3]

En effet, les mares du Ferlo constituent à fois la source d'approvisionnement en eau pour les populations et le bétail tout en étant le gîte larvaire des vecteurs. Dans cette zone, la variabilité intra-saisonnière de la pluviométrie, la dynamique de la végétation et la turbidité des mares temporaires, dont la taille est relativement petite, sont les facteurs principaux qui expliquent la forte concentration des moustiques [18]. Pour comprendre et maîtriser le processus des maladies vectorielles, il est nécessaire de prendre en compte, dans un même environnement, les phénomènes multiples et variés, mais aussi à différentes échelles (temporelle, spatiale et organisationnelle). Les partenaires impliqués dans le projet QWeCI (<http://www.liv.ac.uk/qweci>) au Sénégal (figure 2) fournissent des données provenant des enquêtes de terrain (données observées ou générées par des équipements de mesure) ou des tests de laboratoire (données d'analyse). Le défi à relever par ces équipes de recherche est la maîtrise de l'impact des facteurs environnementaux sur le déclenchement et la propagation de la FVR. Il est donc fondamental de prendre en compte l'ensemble des données des différentes disciplines dans le modèle de données proposé.

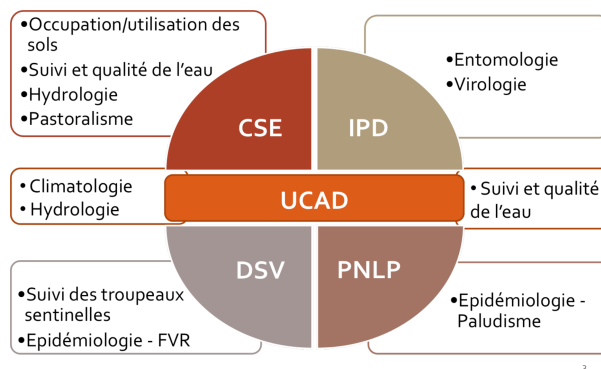


Figure 2. Complexité organisationnelle

3. Matériels et méthodes

3.1. Données manipulées

Les données utilisées pour notre expérimentation sont issues :

- des tests de laboratoires sur la qualité de l'eau de mare ; ils permettent de déterminer les caractéristiques (pH, température, conductivité et turbidité) des eaux prélevées dans les mares étudiées. Ces données couvrent le mois d'Août pour les années 2008 à 2012 avec 72 tests effectués pour les cinq (5) mares ;
- des stations météorologiques (vent, ensoleillement, température, pluviométrie) couvrant les périodes d'Août à Décembre de 2008 à 2012. Ces données sont générées toutes les heures.

3.2. Entrepôt de données

Notre entrepôt de données construit sous le Système de Gestion de Base de Données PostgreSQL repose sur un modèle multidimensionnelle présenté en dans premiers travaux [27].

Nous utilisons le modèle en constellation pour représenter les données. Trois tables de fait ont été identifiées pour la gestion des données environnementales : (1) Climatologie : pour les données générées par les stations météorologiques et les postes pluviométriques ; (2) Hydrologie : pour les mesures du niveau d'eau des mares ; (3) Suivi/Qualité : pour la caractérisation des mares. Les dimensions ont été identifiées en fonction des différents critères d'analyse. On retrouve ainsi les stations/postes, les échelles de mesure et les sites de prélèvement. Ces trois dimensions sont rattachées à la dimension « Mare » qui se retrouve dans l'axe d'analyse spatial suivant la hiérarchie « communauté rurale -> arrondissement -> commune -> département -> région ». La dimension temporelle est gérée en « jour -> mois -> période -> année ».

L'entrepôt de donnée construit est la source de données de nos outils d'expérimentations.

3.3. Méthodologie utilisée : motifs spatiaux temporels

La fouille de données est une étape phare du processus décisionnel qui permet d'extraire des informations, à priori non évidentes, dans un volume élevé de données.

Dans le cadre de ce travail, nous avons choisi la méthode d'analyse des données spatio-temporelles définie par [2] et [1] ; ceci en nous basant sur les travaux de [7]. La dimension spatiale permet de définir la localisation d'un objet dans un espace géographique et la dimension temporelle le rattache à un pas de temps (date, période, etc.).

Dans cette méthode, les auteurs se focalisent sur l'inclusion de la notion de spatialisation dans le processus d'Extraction de Connaissances à partir des Données (ECD). Ils ont proposé une approche basée sur le pré-traitement des données afin d'inclure les caractéristiques spatiales dans les données temporelles. Les séquences spatiales ainsi obtenues sont ensuite utilisées en entrée de l'étape de fouille de données afin d'en extraire des séquences spatialement fréquentes. Ceci est réalisé à l'aide d'un algorithme d'extraction de motifs séquentiels classique [15] afin d'extraire des motifs représentant les évolutions temporelles spatialement fréquentes des zones étudiées.

4. Résultats

La mise en place de notre environnement décisionnel s'appuie sur le processus décrit par [13]. Nos premiers travaux [27] nous ont permis de proposer un modèle de données qui permet (1) d'identifier les différents axes d'analyse; (2) de corréler les différents indicateurs de mesures de qualité [24]; (3) de proposer des vues qui correspondent aux attentes des différentes disciplines impliquées [10]. Il s'agit dans cette seconde phase de proposer des motifs de description et de prédiction des facteurs environnementaux et sanitaires. Ainsi, nous présentons, dans cette section, les premiers résultats de fouille obtenus.

Les jeux de données concernent les éléments climatologiques (température, vent, pluviométrie ...) et les paramètres de qualité d'eau des mares (teneur en acidité, matières en suspens, température ...). Deux expérimentations ont été faites : des données regroupées par mare et des données regroupées par site. Pour le premier jeu de données (mares), la dimension temporelle est très variée; avec un support minimal de 0.9 (au moins 90% des mares remplissent les conditions de la séquence fournie), on obtient 21131 séquences fréquentes (Tableau 1). Dans cet article, nous avons effectué un extrait arbitraire des motifs obtenus.

Tableau 1. Motif "mare"

Motifs	Support
(ce:[42.45;51.30])(tds:[22.65;27.45])(temp:>29.05)(ph:[5.36;5.54])	1 (5/5)
(ph:[5.54;5.66])(temp:[28.35;29.05])	1 (5/5)
(ph:[5.36;5.54])(tds:[22.65;27.45])(ce:[42.45;51.30])(temp:[28.35;29.05])	1 (5/5)

Ces motifs obtenus permettent de certifier que la qualité des eaux des cinq (5) mares (Figure 1) de la zone du Ferlo est très similaire. En effet, ces motifs ont un support de 100%. Ceci signifie que toutes les mares analysées pendant des périodes similaires ont exactement le même comportement. Dès lors, nous jugeons également plus pertinent de confronter les variantes des mares à celles de l'environnement météo-climatique pour maîtriser l'impact de ces derniers sur la qualité des eaux des mares.

Ainsi, en étudiant les mêmes données sur la base des stations et postes pluviométriques, nous avons obtenus des supports très diversifiés qui varient entre 44% (ph, température, matières) et 94% (pour la température) (Tableau 2). Ce jeu de données comporte plus de zones géographiques (localités) mais moins de dates. Les résultats sont composés de 67912 séquences fréquentes pour un support minimal de 0.4 (au moins 40% des localités remplissent les conditions de la séquence).

Tableau 2. Motif "site"

Motifs	Support
(temp:[28.35;29.05])	0,94 (15/16)
(ph:<=5.45)(ce:[42.45;51.30])(tds:[22.65;27.45])	0,75 (12/16)
(ph:(5.45;5.62])(temp:[28.35;29.05])(tds:<=22.65)	0,44 (7/16)

Ces derniers résultats ne fournissent pas une interprétation cohérente. En effet, cette analyse n'apporte pas de réponses fondées sur d'éventuelles corrélations entre les caractéristiques des eaux des mares et celles fournies par les stations météo-automatiques. Cela pourrait se justifier par la faible densité temporelle.

Les données météorologiques confirment que le climat des communautés rurales de cette zone d'étude est très similaire. Dès lors, il conviendrait de se pencher sur d'autres paramètres tels que la qualité des sols pour pouvoir justifier la diversité vectorielle de la zone d'étude.

5. Conclusion et perspectives

Pour rappel, l'objectif de nos travaux de recherche est de proposer un environnement décisionnel décrivant les interactions entre les indicateurs environnementaux et la propagation de la maladie de la Fièvre de la Vallée du Rift. Dans cet article, nous avons présenté quelques extraits des motifs obtenus sur les données de suivi et de qualité de l'eau des mares. Ces expérimentations nous ont permis de confirmer que les mares de la zone du Ferlo présentent les mêmes caractéristiques de qualité d'eau. Mais l'agrégation des données de caractérisation des mares et celles des stations météorologiques ne

fournit pas d'éléments de discussion ; la faible variabilité temporelle pourrait être un facteur influant. Malgré le jeu de données peu élevé, ce retour d'expérience nous conforte dans notre approche d'intégration des attributs spatiaux et temporels pour la compréhension et la maîtrise du risque sanitaire de la FVR. Pour la suite de nos travaux, il est important de pouvoir confronter les résultats obtenus avec les données entomologiques et virologiques pour proposer des motifs de classification et de prédiction. Par ailleurs, nous travaillons sur la détermination de voisinages des objets géo référencés (mares, stations) suivant l'approche polaire et/ou euclidienne. Ainsi, il s'agira d'identifier toutes les corrélations entre les paramètres geo-environnementaux et les données entomologiques et sanitaires en tenant compte de caractères fondamentaux tels que l'espace et le temps. De plus, l'observation de la transhumance des troupeaux devrait apporter des éléments supplémentaires pour des projections spatio-temporelles de l'apparition et de la propagation de la FVR. En fouille de données, les données constituent la matière première des informations sur lesquelles des décisions doivent être prises. Aussi, Les problématiques relatives à la qualité des données sont la cause des principaux échecs dans la mise en place d'un système décisionnel. Pour approfondir notre approche, il serait pertinent de proposer une méthode pour reconstruire les données manquantes, tenant compte de leur proportion et de leur type. Enfin, nous essaierons de proposer une technique pour analyser tous les motifs générés sans faire de choix arbitraire comme c'est le cas dans ce travail.

6. Bibliographie

- [1] Agrawal R., Srikant R., *Mining sequential patterns*. In Philip S. Yu and Arbee L. P. Chen, editors, Proceedings of the Eleventh International Conference on Data Engineering (ICDE), Taipei, Taiwan, p. 3-14. IEEE Computer Society, (1995)
- [2] Alatrasta Salas H., Cernesson F., Bringay S., Azé J., Flouvat F., Semaloui N. et Teisseire M., *Recherche de séquences spatio-temporelles peu contredites dans des données hydrologiques*. Revue des Nouvelles Technologies de l'Information (RNTI), numéro spécial Qualité des Données et des Connaissances/Evaluation des Méthodes d'Extraction des Connaissances dans les Données. Vol. RNTI-E-22, P. 165-188, isbn: 978-2-70568-286-6, (2011)
- [3] Coly A., Le système fluviolacustre du Guiers : étude hydrologique et gestion quantitative intégrée, Phd, UCAD, (1996)
- [4] Diagne F.F., Etude de la Fièvre de la Vallée du Rift chez les ruminants domestiques au Sénégal : enquêtes sérologiques dans la vallée du Fleuve, le Ferlo et la Casamance, Phd, FMPOS/UCAD, p. 29-34, (1992)
- [5] Diallo M., Dynamique comparée des populations de Culicidae à Kédougou (zone soudano-guinéenne) et à Barkédji (zone de savane sahélienne) : conséquences dans la transmission des arbovirus, DEA en Biologie Animale, Université Cheikh Anta Diop de Dakar, (1995)
- [6] Diallo M., Lochouarn L., Ba K., Sall A., Mondo M., Girault L., Mathiot C., *First isolation of the Rift Valley fever virus from Culex poicilipes (Diptera : Culicidae) in nature*, American Journal of Tropical Medicine and Hygiene, vol. 6, no 62, (2000)

- [7] Fabrègue M., Braud A., Bringay S., Le Ber F., Teisseire M., *Extraction de motifs spatio-temporels à différentes échelles avec gestion de relations spatiales qualitatives*, Inforsid 2012, Montpellier (2012)
- [8] Flick R., Bouloy M., *Rift valley fever virus*, Curr Mol Med 5, 827–834, (2005)
- [9] Golfarelli M., Maio D., Rizzi S., *The Dimensional Fact Model: A Conceptual Model for Datawarehouses*. International Journal of Cooperative Information Systems, 7(2-3), p. 215-247, (1998)
- [10] Golfarelli M., Rizzi S., Saltarelli E., WAND, *A Case Tool for Work load Based Design of a Data Mart*. 10th National Convention on Systems Evolution for Databases, pp. 422-426, (2002)
- [11] Greboval M., Facteurs environnementaux influençant la dynamique des vecteurs du virus de la fièvre de la Vallée du rift : conséquences pour la modélisation de la maladie, Thèse de Doctorat Vétérinaire, Ecole Nationale Vétérinaire de Lyon, (2003)
- [12] Hechmati G., Epidémies de grippe : système d'information pour la prise de décision en santé publique, Phd Thesis, Université de Genève, (2004)
- [13] Marakas G.M., *Decision Support Systems in the 21st Century*. Prentice Hall, Upper Saddle River, (2003)
- [14] Meliker J., Sloan C., *Spatio-temporal epidemiology: Principles and opportunities*. Spatial and Spatio-temporal Epidemiology 2, 1–9, (2011)
- [15] Mortazavi-Asl B., Pinto H et Dayal U., *Prefix Span : mining sequential patterns efficiently by prefix-projected pattern growth*. Proceedings 17th International Conference on Data Engineering, pages 215-224, (2000)
- [16] Murty U., Rao M., Misra S., *Prioritization of malaria endemic zones using self-organizing maps in the Manipur state of India*, Informatics for Health and Social Care, vol. 33, no 3, p. 170-178, (2008)
- [17] Ndione J.A., Besancenot J., Lacaux J., Sabatier P., *Environnement et épidémiologie de la Fièvre de la Vallée du Rift (FVR) dans le bassin inférieur du fleuve Sénégal*. Environnement, Risques et Sante 3(2), 176–182, (2003)
- [18] Ndione J.A., Lacaux J.P., Tourre Y., Vignolles C., Fontanaz D., Lafaye M., *Mares temporaires et risques sanitaires au Ferlo : contribution de la télédétection pour l'étude de la Fièvre de la Vallée du Rift entre aout 2003 et janvier 2004*, (2009)
- [19] OMS, *La Fièvre de la Vallée du Rift*. Aide-Memoire 207, Nations Unies, <http://www.who.int/mediacentre/factsheets/fs207/fr/index.html>, (2010)
- [20] Prehaud C., Bouloy M., *La Fièvre de la Vallée du Rift : un modèle d'étude des fièvres hémorragiques virales*, Technical report, Institut Pasteur, (1997)
- [21] Ravat F., Teste O., Zurfliuh G., *Algèbre OLAP et langage graphique*, CoRR abs/1005.0213, (2010)
- [22] Robertson C., T.A. Nelson, Y.C. MacNab, A.B. Lawson, *Review of methods for space-time disease surveillance*. Spatial and Spatio-temporal Epidemiology, 1:105-116, (2010)
- [23] Tran A., Biteau-Coroller F., Guis H., Roger F., *Modélisation des maladies vectorielles*, Epidémiol. et santé anim., 47, 35-51, (2005)
- [24] Trujillo J. C., Luján-Mora S., et Song I., *Applying UML for designing multidimensional databases and OLAP applications*. K. Siau (Ed.), Advanced Topics in Database Research, Vol. 2, Idea Group Publishing, p. 13-36, (2003)
- [25] Wehrle P., *Modele multidimensionnel et OLAP sur architecture de grille*, PhD thesis, Institut national des sciences appliquees de Lyon, (2009)
- [26] Information on <http://www.oie.int/fr/sante-animale-dans-le-monde/maladies-de-la-liste-de-loie-2013/>, (2013)
- [27] Bouba F., Bah A., Cambier C., Ndiaye S., Ndione J.A., *Modèle multidimensionnel sur les interactions environnement-santé : Cas de la Fièvre de la Vallée du Rift au Sénégal*. CNRIA 2012, 2012.