Géostatistique des phénomènes à tendance périodique (dans l'espace-temps)

Serge Antoine Séguret

▶ **To cite this version:**

Serge Antoine Séguret. Géostatistique des phénomènes à tendance périodique (dans l'espace-temps). Mathématiques [math]. École Nationale Supérieure des Mines de Paris, 1991. Français. NNT : 1991ENMP0247 . tel-02174889

HAL Id: tel-02174889
https://hal.science/tel-02174889

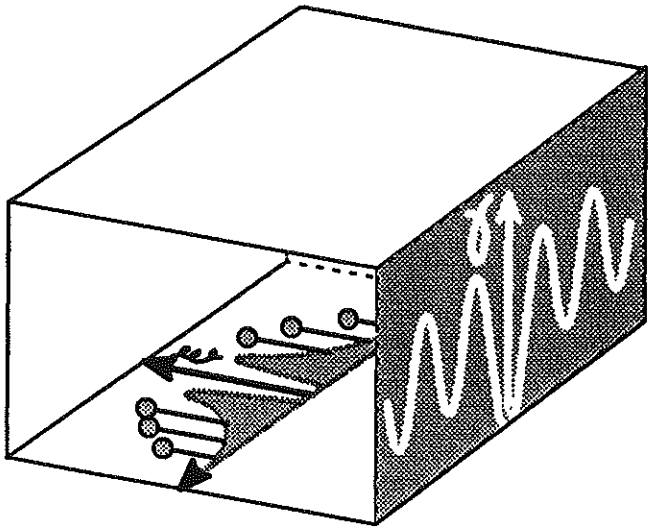
Submitted on 5 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GÉOSTATISTIQUE DES PHÉNOMÈNES À TENDANCE PÉRIODIQUE

(dans l'espace-temps)



THÈSE

présentée à

L'Ecole Nationale Supérieure des Mines de Paris

par

Serge Antoine SÉGURET

pour obtenir le titre de

Docteur en Géostatistique

Soutenue le 22 Février 1991
devant le jury composé de :

M.	G. MATHERON	Président
M.	Ph. HUCHON	Rapporteur
Mlle	Ch. de FOUQUET	Rapporteur
M.	P. CHAUVET	Examinateur
M.	V. RENARD	Examinateur
M.	A. TARANTOLA	Examinateur

Géostatistique des Phénomènes à Tendance Périodique (dans l'espace-temps)

Remerciements		page 3
Résumé de la thèse en Français (Anglais)		page 5 (7)
Introduction		page 9
Chapitre 1	Les Phénomènes à Tendance Périodique (P.T.P.)	page 13
Chapitre 2	Analyse de P.T.P. temporels (perturbations diurnes en magnétisme).	page 25
Chapitre 3	Analyse de P.T.P. spatio-temporels (magnétisme).	page 65
Chapitre 4	Estimation de P.T.P. spatio-temporels (magnétisme)	page 97
Chapitre 5	Autres applications (imagerie)	page 149
Conclusions		page 163
Annexes		page 177
Nomenclature des notations, abréviations et mots-clefs		page 239
Bibliographie et références		page 243
Table des figures		page 247
Table des matières détaillée		page 251

Si je tiens à remercier le Professeur G. Matheron, Directeur du Centre de Géostatistique, c'est non seulement pour l'aide qu'il m'a apportée dans la réalisation de ce travail, mais aussi pour l'intérêt qu'il lui a porté, et surtout pour m'avoir enseigné "l'Éthique Géostatistique", cette ligne de conduite scientifique que quelques phrases ne sauraient définir, mais que peut-être le travail qui suit permettra d'appréhender.

Ma plus grande gratitude ira ensuite à V. Renard, Géophysicien de l'Institut Français de Recherche et d'Exploitation de la Mer (I.F.R.E.MER), qui, non content d'avoir su me poser avec clarté et discernement le problème des perturbations en magnétisme, s'est donné la peine de m'initier aux rites de la Géophysique Marine en m'emmenant sur le Navire Océanographique Jean Charcot, en bordure de la banquise ...
Cette admiration que je lui porte, mêlée d'une amitié qui s'affermit de jour en jour, ont été ma plus grande motivation pour faire aboutir les méthodes élaborées dans cette thèse.

Comment dans ces remerciements ne pas faire suivre un Géophysicien par un autre, de l'École Normale Supérieure cette fois, j'ai nommé Ph. Huchon ? Son vif intérêt pour nos techniques, sa rigueur, ses propos motivants, autant de faits m'ayant permis de surmonter bien des obstacles. Un grand merci à lui.

Et pour ne pas quitter le domaine de ceux que nous appelons les "praticiens" de nos méthodes (mais qui sont bien souvent de remarquables "théoriciens" d'autres méthodes ...) je tiens à remercier A. Tarantola, Physicien de l'Institut de Physique du Globe de Paris, d'avoir bien voulu assister à ma soutenance en la position de membre du jury.

Mon directeur de thèse fut P. Chauvet, chercheur au Centre de Géostatistique (C.G.). Je tiens à le remercier pour avoir assumé son rôle jusqu'à son terme. Il en est de même de C. de Fouquet, chercheur au C.G., dont la méticulosité à décortiquer mon travail fut très instructive.

Je me dois également de remercier A. Galli (Chercheur au C.G.) pour m'avoir suggéré en 1987 de concevoir une "Dérive Externe Trigonométrique" et J. Rivoirard (Chercheur au C.G.), de m'avoir la même année susurré de travailler dans l'espace-temps alors que je n'étais qu'un être misérablement mono-dimensionnel...

Ces considérations techniques me conduisent à transmettre à D. Renard (Chercheur et Informaticien au C.G.) mes plus vifs remerciements pour s'être aimablement prêté à de véritables tortures sur son remarquable logiciel BLUEPACK, ceci afin de me permettre de réaliser bon nombre d'expériences. De plus, ses explications sur le fonctionnement de cet outil, ainsi que ses remarques de fond et de forme sur le manuscrit de ma thèse, m'ont considérablement aidé pour améliorer mon travail.

Et puisque je suis dans le domaine des spécialistes de l'Informatique Géostatistique, je tiens à remercier Y. Touffait, non pour des considérations informatiques, mais pour son soutien actif à certains moments critiques de ma carrière au C.G..
De même, je tiens à transmettre toute ma sympathie à F. Geffroy (Chercheur et Informaticien au C.G.) et le remercier de son inaltérable amabilité.

Outre les personnes précitées, il me faut remercier mes autres collègues du Centre qui, à des titres divers et à maintes occasions, ont su contribuer à

transformer le petit quotidien en une grande aventure : H. Beucher, G. Le Loc'h, C. Daly, H. Wackernagel, L. Hu, D. Jeulin, Ch. Lantuejoul, Ch. Lajaunie, M. Armstrong.

Si j'ai pu vivre cette thèse, c'est surtout grâce à la bienveillance d'une Grande Dame: l'Ecole Nationale Supérieure des Mines de Paris, et à ses responsables auxquels je souhaite rendre hommage.

Si j'ai eu matière à recherche (les données ... les problèmes !) c'est essentiellement grâce :

au Département Géosciences Marines de l'I.F.R.E.MER à Brest,

au Département de Géologie de L'Ecole Normale Supérieure de Paris,

au Centre de Morphologie Mathématique de Fontainebleau.

Merci aux nombreuses personnes contactées qui travaillent dans ces organismes remarquables.

Je tiens également à remercier I. Schmitt et M. Briffe pour leur contribution à la frappe de ce travail, M. Armstrong pour ses traductions et surtout rendre hommage à J. Laurent qui a magistralement réalisé l'intégralité des quelques deux cent quarante illustrations qui émaillent le texte.

Je souhaite dédier ce travail à ...

... mes amis :

V., P., A., F., L., J. Clément et proches,

M. Courtillat et proches,

J.P. Gruet et proches,

J. Boujot, D. Gilleron, E. Larrouy,

... mes Parents, mon Frère Patrick et nos proches,

... ma merveilleuse Compagne C. K..

Quelque part, cette thèse n'appartient qu'à eux.

Résumé de la thèse

L'essentiel de ce travail est consacré à l'étude du magnétisme terrestre et à sa séparation en termes temporel (les perturbations) et spatial (les anomalies). Le lecteur uniquement intéressé par ce problème aura avantage à lire l'article rédigé en Anglais et référencé **[GST-14]**. Il constitue une bonne introduction à cette thèse.
Le comportement quasiment périodique des perturbations temporelles nous a amené à définir et examiner une classe de phénomènes naturels : les Phénomènes à Tendance Périodique (P.T.P.). Ils possèdent cette particularité qu'au niveau de leurs variations, on ne peut définir objectivement une période, étant donné leur caractère fluctuant et imprévisible. Il n'en est pas de même de leurs caractéristiques statistiques (variogramme,spectre) qui, élaborées sur un nombre même restreint de "quasi" périodes (six ou sept pour l'exemple du magnétisme) révèlent une composante périodique, qui plus est constituée de peu de fréquences (deux ou trois dans l'exemple). Ces caractéristiques ont conduit à utiliser des techniques classiques de Géostatistique Linéaire (variogramme, Krigeage) conjointement à des techniques spectrales élémentaires (spectre, Transformée de Fourier discrète (T.F.d.)) et d'élaborer des outils d'estimations qui chevauchent les mondes spatiaux et spectraux : le Krigeage Trigonométrique (K.T.) et l'Analyse Krigeante Trigonométrique (A.K.T.), deux manières d'effectuer le Krigeage de coefficients de Fourier et donc de généraliser la T.F.d..

Outre quelques pages introductives, conclusives et des annexes qui (A) présentent l'opérateur T.F.d., (B) les outils géostatistiques, (C) le Krigeage, (D,E et F) le détail de modélisations, ce travail se présente en cinq chapitres . Ci suit leur contenu.

Le chapitre 1 : les Phénomènes à Tendance Périodique (13 pages), contient une description des P.T.P. et quelques exemples : pollution, température de l'eau, magnétisme, zone d'ouverture océanique, cellules microscopiques.
Si ce caractère répétitif peut se manifester dans le plan ou un espace de dimension quelconque, on note que c'est bien souvent dans le temps que cette propriété intervient. La régionalisation usuelle du P.T.P. est donc le champ (spatial) de séries temporelles.

Le chapitre 2 : analyse de P.T.P. temporels (42 pages) est consacré à l'analyse de trois séries temporelles de perturbations du magnétisme terrestre: les séries Kakioka, Huancayo et Mohns.
Pour la série Kakioka, le fait de disposer de soixante jours de mesure par pas de une minute a permis d'appréhender très finement les fluctuations diurnes (période 24 heures) et quelque peu les perturbations lunaires (période 28 jours). L'utilisation conjointe de la T.F.d. et du variogramme a conduit à une modélisation très précise du variogramme, à l'échelle de la minute, pour les trois premiers jours.
Pour la série Huancayo, nous ne disposions que de six jours de mesure, ce qui a conduit à deux autres modes d'ajustement : l'un procède d'une détection manuelle des fréquences et énergies, l'autre utilise uniquement l'opérateur T.F.d.. Les résultats s'avèrent aussi bons que ceux de la série Kakioka.
Pour la série Mohns, le fait de ne disposer que de trois jours de mesure, conjointement à la localisation boréale de l'appareil de mesure (où on sait que les perturbations oscillent faiblement) n'ont pas permis de détecter avec précision une quelconque composante périodique dans le variogramme

Le chapitre 3 : analyse de P.T.P. spatio-temporels (33 pages) s'attache à l'analyse et la modélisation de trois jeux de données spatio-temporelles de magnétisme : les campagnes océanographiques Krakatao, Seaperc et Mohns. Si le détail des modélisations figure en annexe D, E et F, c'est dans le but de mieux mettre en avant le principal problème rencontré pour utiliser le K.T. sur ce type de données: à savoir que l'échantillonnage du bateau est quasiment "coplanaire" dans l'espace-temps, ce qui limite singulièrement notre latitude pour appréhender les oscillations temporelles à partir du signal spatio-temporel magnétisme. En effet, le choix de modéliser le magnétisme par la somme d'une Fonction Aléatoire Intrinsèque d'ordre k (F.A.I.k en abrégé) pour les anomalies spatiales avec une série de Fourier (pour les perturbations temporelles) nécessite un échantillonnage spatial isotrope et un échantillonnage temporel dont le pas doit être adapté à la période 24 heures (celle des perturbations). Or, d'une part les données sont sur des profils parallèles (ce qui pose des problèmes pour inférer les paramètres de la F.A.I.k) et d'autre part, l'espace et le temps sont liés par la vitesse du bateau, de sorte que lorsqu'on tente d'améliorer l'échantillonnage spatial (en choisissant des données particulières), on détruit bien souvent l'échantillonnage temporel, et réciproquement.
Malgré cette incontournable limitation, les campagnes Krakatao et Seaperc ont permis de détecter des oscillations temporelles de période 24 heures et 12 heures et d'obtenir des modèles sans trop de difficulté. Par contre, il n'en est pas de même pour la campagne Mohns où le rapport signal sur bruit est bien moins favorable. Il résulte de ces trois cas d'étude que nos techniques de détection et d'estimation dans ce contexte d'échantillonnage, ne se prêtent qu'à des P.T.P. possédant des ondulations temporelles à forte énergie comparativement aux anomalies spatiales.

Le chapitre 4 : estimation de P.T.P. spatio-temporels (53 pages) utilise les modèles inférés dans le chapitre 3 pour filtrer les perturbations temporelles et cartographier les anomalies spatiales du magnétisme. L'outil utilisé est le K.T.. Les résultats sont bons pour Seaperc et Krakatao, mais désastreux pour Mohns, ce que les résultats du chapitre 3 laissaient prévoir. D'une manière générale, certaines améliorations informatiques devraient permettre d'obtenir un estimateur temporel moins bruité qu'il ne l'est actuellement.

Le chapitre 5 : autres applications (15 pages) est consacré à l'imagerie, et ceci à travers deux problèmes.
La première application est homologue du problème des perturbations en magnétisme : l'image initiale (le visage de Gandhi) est considérée comme issue d'un balayage horizontal. Elle est bruitée par une sinusoïde de phase et d'amplitude inconnues. On utilise ensuite le K.T. pour filtrer le bruit et restaurer l'image. Cette expérience est présentée en complément de celles du magnétisme et montre que lorsque nous sommes dans un cadre optimal (pas de problème d'échantillonnage, image modélisée par une Fonction Aléatoire stationnaire), le K.T. est très efficace.
La deuxième application est une étude comparative entre Analyse Krigeante classique (A.K.) et Analyse Krigeante Trigonométrique pour estimer les contours du visage de Gandhi (que l'on associe à la période 8 pixels). Par A.K., on capte toutes les fréquences spatiales comprises entre zéro et la période ciblée (huit pixels dans l'exemple) alors que par A.K.T., on ne capte que la période ciblée (huit pixels), ce qui se manifeste par des contours estimés plus fins.
Cette dernière expérience illustre l'intérêt de parfois décomposer une covariance ou un variogramme en série de Fourier.

Abstract of the thesis

This work is concerned mainly with the earth's magnetism and its division in terms of time perturbations and space anomalies. The reader who is only interested in this problem should read article **[GST-14]**. It is in English.
The almost periodic behavior of the time perturbations has led us to define and examine a class of natural phenomena called Periodic Tendancy Phenomena (P.T.P.). A period cannot be defined objectively for these because of their varying and unpredictable nature. This is not true for its statistical characteristics (variogram, spectrum) which when calculated from even a short number of "almost periods" (six or seven in the example of magnetism) reveal a component quite periodic constituted of a few frequencies (here two or three). These characteristics lead us to develop two estimators based on classical linear geostatistical methods (variogram, kriging) together with elementary spectral techniques (spectrum, discrete Fourier Transforms (d.F.T.)) : the Trigonometric Kriging (T.K.) and the Trigonometric Kriging Analysis (T.K.A.). These are two ways for kriging Fourier series and then for generalizing the d.F.T. operator.

In addition to the introduction, the conclusion and six appendices this work is divided into five chapters. Their content is as follows.

Chapter 1 : the Periodic Tendancy Phenomena (13 pages) This contains a description of P.T.P. and a few examples taken from the fields of pollution, temperature, magnetism, continental drift, microscopic cells. Although this characteristic of being (almost) periodic can exist in a plane or in a space of any dimension, we notice that it is more common in time. So the usual space for the P.T.P. is (spatial) fields of temporal series.

Chapter 2 : analysis of temporal P.T.P. (42 pages) The analysis of three temporal series of magnetic perturbations for the regions of Kakioka, Huancayo and Mohns series is presented in this chapter.
For the Kakioka series, as we had readings every minute for a period of sixty days we were able to quantify the energies in the diurnal oscillations (24 hour period) precisely and to see the influence of the moon (28 day period). Using jointly the d.F.T. operator and variogram gave a very precise model of the covariance, for first three days at the scale of one minute.
For the Huancayo series, having only the data for six days led us to other techniques for fitting the variogram : one by estimating frequencies and energies "manually" , an another using only d.F.T. operator. The results obtained are as precise as those obtained for Kakioka series.
For the Mohns series, having only readings for three days, as well as the spatial location of the series (north pole where the periodic component of the variations is low compared to magnetic storms and spatial anomalies), enabled us to quantify energies and frequencies.

Chapter 3 : analysis of spatial-temporal P.T.P. (33 pages) The analyses of three sets of spatial-temporal data of magnetism : the oceanographic campains Krakatao, Seaperc and Mohns are reviewed in this chapter. The details of the associated modellings are given in the appendices D, E and F. This was done in order to emphasize the main problem encountered when using T.K.which is the fact that samples taken by the boat are nearly coplanar in the space-time. Consequently it is difficult to distinguish temporal oscillations from the space-time signal. Indeed, if an Intrinsic Random Fonction of order k (I.R.F.-k) is used to model the spatial anomalies, an isotropic sampling in space is needed

whereas using a Fourier series for the temporal perturbations we need a sample–step adapted to the 24 hour period. But the shipboard sampling gives us data on parallel profiles (which gives problems when fitting an I.R.F.–k model). The fact that space and time are connected by the boat's speed implies that changing the spatial sampling (by choosing to use specific data in order to have an isotropic spatial sampling) would destroy the temporal sampling; and conversely.
Despite these difficulties, we obtain precise models for the Krakatao and Seaperc campains. For the Mohns's one, it was impossible to obtain an objective model because of the small amplitude oscillations in time.
The conclusion is that for this type of sampling (oceanographic campains), our techniques must be used only if the P.T.P. has temporal oscillations with high energy compared to that of the spatial ones.

Chapter 4 : estimating spatial–temporal P.T.P. (53 pages) This uses the model presented in chapter 3 to filter the diurnal variations (in time) and to map the spatial anomalies. The results are quite satisfactory for the Krakatao and Seaperc campaigns, but are not for the Mohns's campaign, as could be predicted from chapter 3.
Some computational improvments are proposed to obtain a more regular temporal estimator .

Chapter 5 : other applications (15 pages) Two case studies on image analysis are presented here.
The first application can be compared to the problem of filtering temporal perturbations of the magnetism. The initial image (Gandhi's face) is considered as obtained by an horizontal sweep (a T.V. image) to which a sinusoid with unknown amplitude and phase was added. We used T.K. to filter this noise. This experiment is complementary to those on magnetism because it shows that when we do not have sampling problems, the T.K. is a very efficient method.
The second application is a comparative study between classical Kriging Analysis (K.A.) and Trigonometric Kriging Analysis (T.K.A.) to estimate the contour lines of Gandhi's face (we associate to 8 pixel period). By using K.A., we obtain all the periods including those in the 0 to 8 interval, and with T.K.A., we just obtained the 8 pixel energies. In other words, the contour lines obtained by T.K.A. are finer than those obtained by K.A..
This shows the usefulness of splitting a covariance or variogram (experimental or not) into a Fourier series.

At the end of the thesis are six appendices which presents the d.F.T. operator, some geostatistical tools, the kriging, and give details on how to find models on actual data.

INTRODUCTION

Le filtrage des perturbations diurnes en géophysique marine page 11

Les Phénomènes à Tendance Périodique page 11

La dualité spatial-spectral page 11

Organisation du document page 12

Introduction

Le filtrage des perturbations diurnes en géophysique marine

L'essentiel de ce travail consiste en un descriptif de la manière dont nous résolvons le problème suivant :

> Un bateau se déplace sur une zone d'étude, entraînant derrière lui un magnétomètre dans le but de mesurer les anomalies du fond marin par rapport au champ de référence. Au fil du temps, et du fait d'interactions avec le vent solaire, le niveau de référence du champ magnétique varie.
> Ces variations ne sont pas quelconques mais à forte tendance périodique puisque leur origine est la rotation de la terre sur elle-même et autour du soleil.
> Le problème est que ces fluctuations nuisent à la cartographie du magnétisme spatial y(x) puisque la mesure est spatio-temporelle (notons la z(x,t)).
> Il en résulte la nécessité d'effectuer un filtrage des fluctuations temporelles (notées D(t)) dont la caractéristique essentielle est d'être "approximativement" périodique.

Pour résoudre ce problème de filtrage, un outil fut créé : le Krigeage Trigonométrique (KT). Il offre la particularité de s'appuyer sur une modélisation des perturbations par une série de Fourier temporelle (à coefficients inconnus mais de fréquences connues) et sur une modélisation de l'invariant spatial par une Fonction Aléatoire Intrinsèque d'ordre k (FAI-k).

La mise en oeuvre de ce modèle fut un demi succès en ce sens que si parfois l'estimation des fluctuations temporelles était correcte, celle-ci était bien souvent entachée d'instabilités numériques dont la cause essentielle est liée au mode d'échantillonnage des données : une information quasiment en continu sur des profils relativement éloignés les uns des autres.

Les développements de ces techniques, l'analyse des problèmes rencontrés, leur mise en oeuvre sur plusieurs jeux de données, représentent autant de facettes du premier volet de ce travail : le filtrage des perturbations temporelles du magnétisme spatio-temporel.

Les Phénomènes à Tendance Périodique

Parallèlement à ces investigations, nous fûmes confrontés à d'autres phénomènes naturels qui possédaient cette caractéristique d'une "tendance" à se répéter identiques à eux même, et ceci non nécessairement dans le temps.

Ceci explique la classe que nous avons créée, celle de "Phénomènes à Tendance Périodique" (PTP). Bien que différents en de nombreux points du magnétisme spatio-temporel, les PTP dont nous parlerons devraient pouvoir être analysés et modélisés par des techniques similaires.

C'est là le deuxième volet de ce travail : une tentative de généralisation de techniques d'analyse, de modélisation et d'estimation.

La dualité spatial-spectral

Cette tendance à l'auto répétition nous a invariablement conduit vers des techniques d'analyse fréquentielle, en l'occurrence les Transformées de Fourier Discrète (TFD), et, d'une manière plus générale, nous a confronté au choix de techniques spatiales (ajustement de covariances par exemple) ou de techniques spectrales (analyse de spectres par exemple).

Or, si un PTP possède une tendance à l'auto-répétition, ce qui signifie l'émergence de fréquences particulières dans son spectre, le PTP n'en demeure pas moins très fluctuant.

Ceci explique qu'un recours aux Fonctions Aléatoires (FA) se soit avéré fort utile, et de ce fait à des techniques de Géostatistique qui peuvent être qualifiées de "spatiales" puisque basées sur l'utilisation de covariances et variogrammes.

C'est ainsi que nous nous sommes retrouvé à la frontière de deux mondes méthodologiques (spatial et spectral), empruntant à l'un et à l'autre les outils adaptés : la TFD pour la composante périodique du PTP et les fonctions variogramme ou covariance généralisée pour la composante dite non périodique.

Il en résulte que si bien souvent ces deux mondes méthodologiques peuvent apparaître comme duaux, ils seront dans notre cas parfaitement complémentaires.

C'est cette complémentarité qui, vécue expérimentalement, constitue le troisième volet de notre travail.

Organisation du document

Cette thèse est constituée de cinq chapitres et d'annexes :

Chapitre 1	: Les phénomènes à Tendance Périodique (PTP)
Chapitre 2	: Analyse de PTP temporels (perturbations diurnes du magnétisme)
Chapitre 3	: Analyse de PTP spatio-temporels (magnétisme)
Chapitre 4	: Estimation de PTP spatio-temporels (magnétisme)
Chapitre 5	: Autres applications (imagerie)

En ce qui concerne les domaines d'application, on peut regrouper ces chapitres en trois sous-ensembles :

- Le chapitre 1 qui donne plusieurs exemples de PTP, et une définition intuitive de ceux-ci.
- Les chapitres 2, 3 et 4 qui contiennent des expérimentations sur la seule variable magnétisme, pour des régionalisations temporelles et spatio-temporelles.
- Le chapitre 5 qui offre l'occasion de présenter succinctement des applications de nos techniques à l'imagerie.

En ce qui concerne la méthodologie, les chapitres 2 et 3 sont à dissocier des chapitres 4 et 5 puisque le chapitre 2 (resp. 3) s'attache à l'analyse de séries temporelles (resp. de champs de séries temporelles) alors que les chapitres 4 et 5 concernent des estimations et des filtrages.

Une conclusion d'une dizaine de pages permettra de dégager la quintessence des nombreuses expériences contenues dans les chapitres précédents, ainsi que de proposer des améliorations aux techniques élaborées dans cette thèse.

On trouvera en annexe :

- une présentation succincte de la TFD,
- une présentation succincte de la fonction variogramme, des FAI-k et de la manière dont est conduite l'inférence de ces modèles ,
- une présentation du (des) Krigeage(s).
- le détail de l'inférence des modèles associés aux campagnes océanographiques Krakatao, Mohns et Seaperc qui sont le support expérimental des chapitres 3 et 4.

CHAPITRE 1

LES PHÉNOMÈNES À TENDANCE PÉRIODIQUE (P. T. P.)

RÉSUMÉ

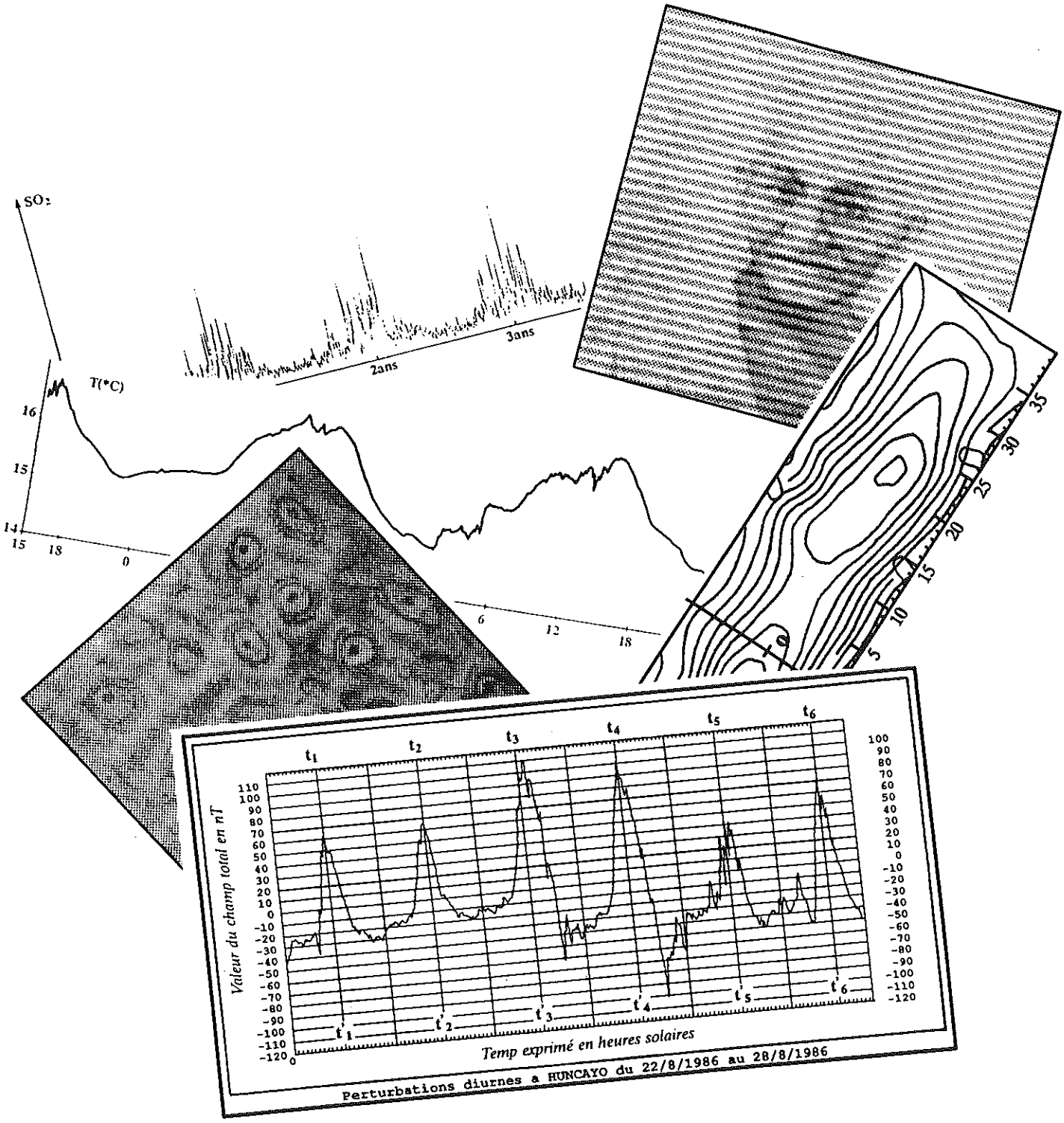
Ce chapitre est consacré à la classe des phénomènes naturels appelés PTP.
Ces phénomènes y sont décrits d'une manière intuitive en des termes spatiaux, spectraux puis dans le langage de la Théorie du Signal.
Ensuite, nous en montrons quelques exemples dont les champs de séries temporelles à caractère périodique dans le temps (pollution, température de l'eau, magnétisme) et certains autres dont cette tendance à l'auto-répétition se manifeste dans l'espace (zone d'ouverture océanique, cellules microscopiques, ...).

Les éléments suivants se dégagent :

- Un PTP est une variable régionalisée de dimension quelconque qui particularise certaines fréquences dans son spectre d'énergie, tout en conservant un caractère suffisamment imprévisible pour qu'un recours au formalisme des Fonctions Aléatoires s'avère utile.
- Pour modéliser un PTP, il faudra considérer ces fréquences privilégiées (en utilisant des techniques spectrales), tout en tenant compte de l'aspect aléatoire du phénomène (a l'aide de techniques spatiales).

1.1 Description page 15

1.2 Exemples page 17


SO2
2ans
3ans
T(°C)
16
15
14
15
18
0
6
12
18
5
10
15
20
25
30
35
t1
t2
t3
t4
t5
t6
t'1
t'2
t'3
t'4
t'5
t'6
Valeur du champ total en nT
Temp exprimé en heures solaires
Perturbations diurnes a HUNCAYO du 22/8/1986 au 28/8/1986

Chapitre 1

LES PHÉNOMÈNES À TENDANCE PÉRIODIQUE (PTP)

Dans la section 1.1, nous présentons l'ensemble des variables régionalisées (VR) que l'on souhaite appréhender dans cette thèse : les Phénomènes à Tendance Périodique, dont la particularité est d'osciller, tout en gardant un comportement suffisamment imprévisible pour que le recours aux Fonctions Aléatoires s'avère indispensable.

Bien que la seule Variable Régionalisée (VR) sur laquelle nous travaillerons dans le détail soit le phénomène des perturbations diurnes en magnétisme terrestre, les enseignements de nos analyses se prêtent à bien d'autres phénomènes.
C'est pourquoi nous présentons dans la section 1.2 quelques VR qui s'apparentent au magnétisme puisque répondant à la définition de la classe de phénomènes que nous construisons.

1 .1 Description

Dans ce qui suit, nous présentons d'une manière intuitive ce que représente le Phénomène à Tendance Périodique (PTP). Dans ce but, nous nous exprimerons successivement en termes spatiaux (section 1.1.1), en termes spectraux (section 1.1.2) puis dans le langage de la Théorie du Signal (section 1.1.3).

1 .1 .1 En termes spatiaux

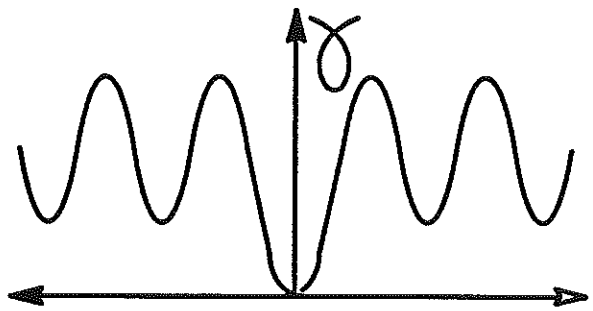
On conviendra d'appeler Phénomène à Tendance Périodique (PTP) , tout phénomène régionalisé dans un espace à n dimensions, et dont on observe une "tendance" à se répéter "quasiment" égal à lui-même dans une ou plusieurs de ces n dimensions.
Plus explicitement, soit z(x) un tel phénomène, avec :

$x = (x_1, ..., x_n)$ le vecteur des coordonnées.

Nous dirons que z(.) est un P.T.P s'il existe $T = (T_1, ..., T_n)$, tel que pour tout x, $z(x + T)$ soit "très peu différent" de z(x).

On désignera T par la période principale du PTP. Dans T, certains T_i sont non nuls. Nous appellerons "espace-période" le sous espace correspondant à l'ensemble des dimensions pour lesquelles les T_i sont non nuls.

Plaçons nous dans l'espace-période du PTP, et examinons son variogramme expérimental selon l'un quelconque des axes de périodicité (disons T_i).
Ce variogramme (notons le $\gamma_{/T_i}(.)$) ne sera pas quelconque, car il devra posséder une composante périodique, ie avoir un comportement comparable à celui du schéma suivant:



Cette fonction sera alors modélisée de la manière suivante :

$$\gamma_{/T_i}(h_i) = \sigma_i(h_i) + \sum_{k=1}^{NFreq} \xi_k(1 - \cos(2\pi f_k h_i)) \qquad (1.1)$$

avec :

$\sigma_i(.)$: un modèle de variogramme stationnaire .

ξ_k : énergie associée à la fréquence f_k.

Examinons chacun de ces points.

Espace-période :

Si nous avons été amené à parler d'espace-période, c'est parce que la majorité des PTP auxquels nous avons à faire ne manifeste de périodicité que dans un sous-espace de l'espace initial de régionalisation. Il en est ainsi (par exemple) du magnétisme spatio-temporel qui oscille dans le temps ou la pollution atmosphérique et la température à la surface de la mer. (Ces exemples seront détaillés ultérieurement).

Dans ces trois cas, la dimension de l'espace période est 1, c'est l'axe des temps . Mais cet espace-période peut également être confondu avec l'espace initial. (voir l'exemple des cellules de polypiers).

Période principale du PTP :

Il reflétera bien entendu les périodes principales moyennes qui sont contenues dans la VR. Ainsi, dans le cas des perturbations diurnes en magnétisme, si leur analyse révèle des pseudo-périodes qui varient entre 21 ^{h} et 25 ^{h} , un examen de leur variogramme calculé (dans le temps) sur un champ même restreint révèlera une période de 24 ^{h}.

En choisissant de caractériser ce PTP par un résultat statistique portant sur le variogramme (et non sur la VR elle-même) on évite l'écueil du calcul (à partir de $z(.)$) de cette "quasi" période, et on se borne à constater le résultat statistique.

Modèle $\sum_{k=1}^{Nfreq} \xi_k(1 - \cos(2\pi f_k h_i))$:

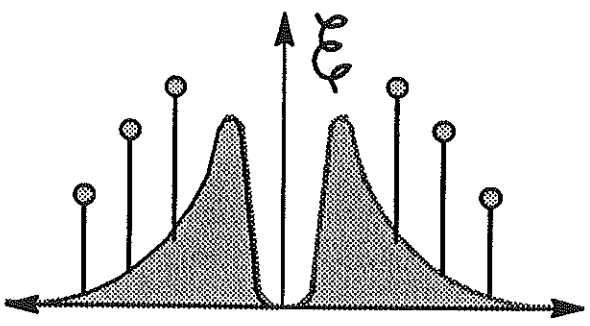
Il est clair que pour h_i variant dans un domaine borné, il serait bien souvent possible d'ajuster $\gamma_{/T_i}$ par Série de Fourier (ce que nous ne manquerons pas de faire dans la suite). Et de tels ajustements s'avèrent possibles même pour des phénomènes n'ayant aucune tendance à l'"auto-répétition" .

Notre propos est en fait (par le biais de la formule 1.1) de dire que le recours à un modèle en cosinus est inévitable, du fait de cette tendance à l'auto-répétition.

Il convient de préciser qu'à l'inverse d'une décomposition en Série de Fourier, la formule 1.1 devra être mise en oeuvre avec un nombre restreint de fréquences (pour les perturbations diurnes en magnétisme, 2 fréquences s'avéreront largement suffisantes pour la modélisation 1.1).

1 .1 .2 <u>En termes spectraux</u>

En termes spectraux, les PTP se caractérisent par l'émergence, dans le spectre d'énergie, de certaines fréquences qui possèdent une énergie plus importante que d'autres . Ci suit le schéma d'un tel spectre :



Ces fréquences (qui rejaillissent dans les oscillations du variogramme) méritent alors dans notre modélisation un statut à part, ce qui impose le recours à un modèle du type 1.1.

1 .1 .3 En termes de Théorie du Signal

Pour qu'un émetteur puisse transmettre loin une information essentiellement composée de hautes fréquences, il est nécessaire d'effectuer au préalable une "modulation" de ce signal. Il en existe de plusieurs types (modulation d'amplitude, de fréquence, etc.....) et mon propos concerne plus particulièrement la "modulation à porteuse sinusoïdale" : celle qui consiste à ajouter au signal initial (par exemple votre émission préférée) une "porteuse" basse-fréquence qui "portera loin". Lorsque vous réglez votre récepteur pour capter l'émission choisie, vous ne faites que régler votre "démodulateur" sur la porteuse (connue) qui est affectée à l'émission choisie. C'est ainsi que l'identification (et l'écoute) deviennent possibles.

D'un point de vue méthodologique, cette porteuse sinusoïdale, par son existence et sa régularité, aura permis d'appréhender bien d'autres fréquences (l'émission favorite tant attendue ...).

Cette image est une autre manière de présenter les PTP, celle de VR qui sont "naturellement" modulées. En réglant notre analyseur de signal (ie : krigeage, cf annexe C) sur certaines fréquences (ie : Krigeage Trigonométrique, cf annexe C), nous devenons capables de démoduler le signal (ie : estimer une composante, inférer une covariance, etc.....).

Si les conditions expérimentales permettent d'appréhender la porteuse, (ie : l'antenne réceptrice est-elle suffisamment grande ?) ces périodes constituent une réalité essentielle qu'il serait regrettable d'omettre de nos modélisations.

1 .2 Exemples

Dans ce qui suit, nous proposons quelques exemples de PTP dont l'espace-période possède une ou deux dimensions. Mis à part le magnétisme terrestre, ces variables régionalisées n'ont pas fait l'objet d'une étude approfondie à l'aide des outils utilisés dans cette thèse. On se bornera donc à les décrire.

Par clarté, nous avons choisi de décomposer ces exemples en deux classes :

- les champs de séries temporelles, pour lesquels l'espace période est le temps,
- d'autres exemples dans lesquels l'espace-période est l'espace à une ou deux dimensions.

1 .2 .1 Les champs de séries temporelles

Les trois exemples de P.T.P qui suivent concernent chacun un ensemble de séries temporelles régionalisées dans l'espace. Examinées chacune séparément ces séries possèdent cette "tendance" à l'auto-répétition dont il était question précédemment.

A PTP fixé, il existera donc trois problèmes à résoudre quant à ces phénomènes :

- Quelles sont les caractéristiques spectrales de ces séries ?
- Comment varient ces caractéristiques d'une série à l'autre ?
- Comment, dans le but d'une estimation ou d'un filtrage, intégrer ces caractéristiques dans un modèle spatio-temporel ?

Examinons ces exemples.

Pollution :

La courbe de la figure 1.1 est l'une des trente séries temporelles enregistrées en divers points de la ville de Berlin du 1/1/1976 au 8/7/1981. Elle est extraite d'un rapport de C. KIRCH (ref [GΦ-5]). Elle montre l'évolution dans le temps du dioxyde de soufre (unité : 10 μg par m^3).

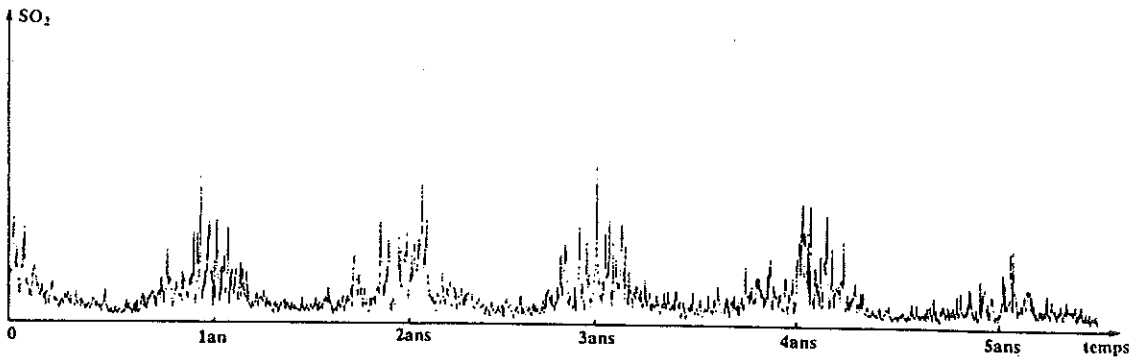


Figure 1.1 : Evolution dans le temps du taux de dioxyde de soufre sur la ville de Berlin (d'après C. Kirch, juin 1982)

D'un point de vue PTP, les caractéristiques sont :

- Espace de la VR : trois dimensions (plan et temps).
- Espace-période : le temps .
- Période principale : T = 365 jours.
- Omniprésence de cette période T sur les 30 séries temporelles.
- Pas de déphasage entre les séries.
- Amplitudes fortement fluctuantes (à temps fixé), d'une série à l'autre.

Une particularité spécifique au SO_2 est à noter : il y a un lien entre l'instant de la mesure et l'amplitude de la VR. On note en effet que les fluctuations sont plus importantes sur les crêtes des oscillations que sur leurs creux. Il en résulte que les algorithmes que nous mettrons au point pour le magnétisme ne devraient pas pouvoir s'appliquer directement à cette VR.

Température de l'eau :

La courbe de la figure 1.2 montre l'évolution dans le temps de la température de l'eau mesurée à une profondeur de 1 mètre, au point fixé 31º 07 N, 9º 53 W (littoral marocain). Elle est citée dans les travaux de G. GOHIN (ref. [GΦ-1]).

D'un point de vue PTP, les caractéristiques sont :

- Espace de la VR : trois dimensions (plan et temps).
- Espace-période : le temps.
- Période principale : T = 24 heures.
- Fortes fluctuations d'une implantation spatiale à une autre et d'une période à une autre (influence des nuages).

L'une des caractéristiques du problème est liée à l'échantillonnage de ce type de mesures lorsqu'elles sont effectuées par un capteur traîné par un bateau. Dans ce cas, le champ de séries temporelles se réduit à une seule mesure (dans le temps) pour une implantation spatiale donnée (sauf si le bateau recoupe sa propre route). La circonstance est d'ailleurs la même que pour le magnétisme en Géophysique marine.

Si la zone d'étude s'étend beaucoup en longitude (plus de 500 km) on peut observer des déphasages entre les différentes séries temporelles.
(A l'équateur , environ 20 minutes pour 500 km).

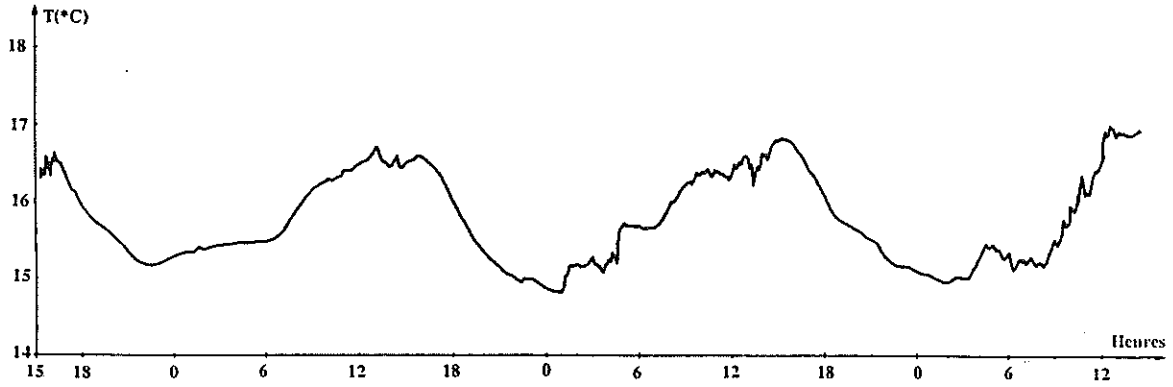


Figure 1.2 : Evolution dans le temps de la température de l'eau. (mesures prises du 05 au 12-08-1973, d'après J.P. Le Saos et J. Talarmin)

En toute généralité, la dimension de la VR est de 4 puisque cet effet diurne s'atténue avec la profondeur de la mesure (atténuation totale à 50 mètres de profondeur).
L'atténuation elle-même (en fonction de la profondeur) dépend de la localisation spatiale (x, y) et du temps, ce qui complique singulièrement le problème.

Magnétisme et perturbations diurnes :

La figure 1.3 contient 6 jours de mesures du magnétisme (à raison d'une mesure par minute) à l'observatoire de Huancayo. Nous aurons longuement l'occasion d'analyser ce phénomène, puisque la majorité des expériences contenues dans cette thèse porte sur cette VR.

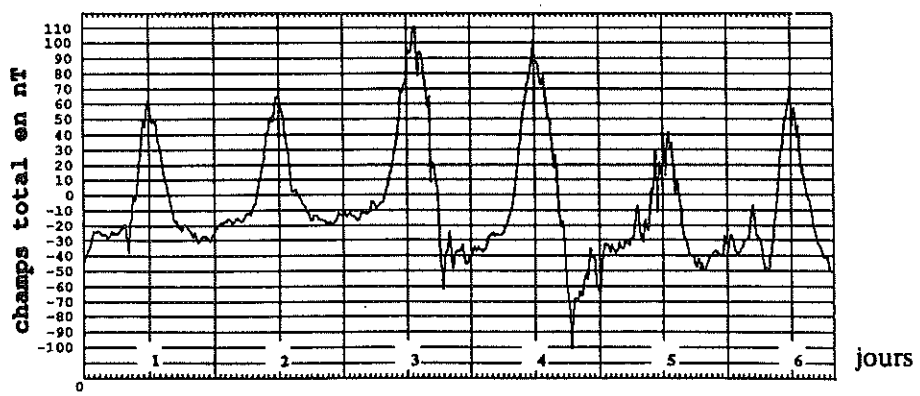


Figure 1.3 : Evolution dans le temps du magnétisme terrestre (mesures prises du 22 au 28-08-1986, observatoire de Huancayo)

D'un point de vue PTP, les caractéristiques sont :

- Espace de la VR : trois dimensions (plan et temps).
- Espace-période : le temps.
- Période principale : 24 heures.

- Faibles fluctuations d'une implantation à l'autre.
- A (x, y) fixé, fortes fluctuations d'une période à l'autre.
- Comme pour l'exemple de la température, on note un déphasage de $\simeq$ 20 minutes entre 2 séries temporelles distantes de 500 km (à l'équateur et pour un déplacement à latitude constante).
- La structure périodique se perd lorsqu'on se déplace de l'équateur vers les pôles.

L'une des caractéristiques du problème est liée à l'échantillonnage spatio-temporel qui est effectué de ce champ de séries temporelles, et qui réduit à un seul temps chaque mesure d'une implantation spatiale donnée.

Conclusions sur les champs de séries temporelles :

Ces trois exemples possèdent en commun le fait que leur espace-période est le temps. Les oscillations (dans le temps) qu'on observe y ont toutes un phénomène déterministe pour origine : la rotation de la terre sur elle-même (cas de la température et du magnétisme) ou la rotation de la terre autour du soleil (exemple de la pollution).

Déjà à ce niveau d'analyse, l'échelle temporelle d'observation de la VR interviendra dans notre conceptualisation du problème puisque, par exemple :

- L'observation du magnétisme sur plusieurs mois montrerait les perturbations lunaires (période : 28 jours). Nous aurons l'occasion d'en faire le constat.
- Si la pollution était mesurée toutes les heures, on verrait apparaître une période de 24 ^{h} également.

Outre cette échelle d'observation (qui est imposée), l'échantillonnage des données conditionne lui aussi la modélisation puisque :

- Dans l'exemple de la pollution (où on dispose de nombreuses séries temporelles), il sera nécessaire d'analyser les variations d'une série à une autre.
- Dans les exemples de la température et du magnétisme (où chaque série n'est connue qu'en un point-temps), une hypothèse de stationnarité (dans l'espace) de la série temporelle s'avèrera nécessaire.

Ces trois faits expérimentaux :

- L'origine déterministe de la composante périodique.
- L'échelle d'observation.
- Le mode d'échantillonnage .

conditionnent d'ores et déjà la modélisation puisque :

- Le premier point suggère d'identifier l'oscillation principale à une sinusoïde,
- Le second point conditionne le choix de la fréquence principale associée au modèle,
- Le troisième point implique d'éventuelles hypothèses de stationnarité (spatiale) des séries (temporelles).

Mais nos séries temporelles sont loin d'être identifiables à des sinusoïdes, car entre la cause (déterministe) et l'effet (imprévisible analytiquement), il existe tout un monde d'interactions peu connues qui engendrent les VR que nous mesurons.

Il en résulte qu'il faudra soumettre ces premiers choix aux contraintes suivantes :

- Identifier l'oscillation principale à une sinusoïde sur un domaine temporel restreint à un segment de période.
- Associer d'autres fréquences annexes à la fréquence principale.
- Déterminer dans quel domaine spatial, l'éventuelle hypothèse de stationnarité est possible.

Mais de tels choix ne peuvent être arbitraires puisqu'ils dépendent de la VR considérée. C'est pourquoi, il faudra avant tout bien "connaître" nos VR.

1 .2 .2 Autres exemples

On se propose maintenant de donner deux autres exemples de PTP.

Cellules microscopiques (image "polypiers") :

La figure 1.4 contient une image de 512x512 pixels qui représente une photographie de cellules de polypiers examinées au microscope.
Cette image nous a aimablement été fournie par J. C. Klein et J. Serra (réf [I–4]).

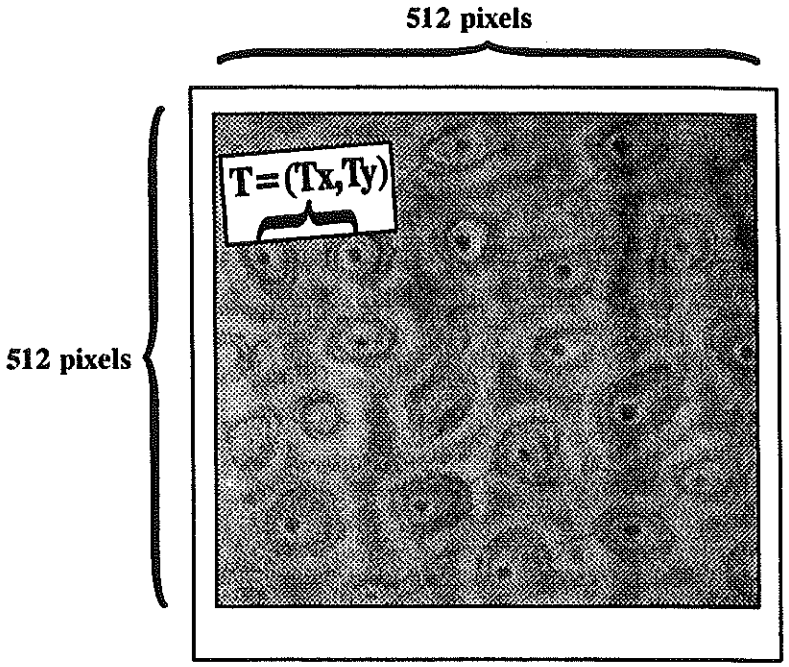


Figure 1.4 : Cellules microscopiques (image "polypiers")

Bien que ce soit sensiblement la même structure (une cellule) qui se répète dans le champ, les formes variables de celle–ci et le caractère imprévisible de son implantation exacte, peuvent conduire à considérer une telle image comme un PTP

Ses caractéristiques sont:

- Espace de la VR : deux dimensions (le plan).
- Espace–période : le plan.
- Période principale : le couple (T_x,T_y) avec T_x, (respt. T_y) distance moyenne en x (respt. en y) entre les cellules.

Zone d'ouverture océanique :

Si la théorie de la tectonique des Plaques a considérablement progressé ces dernières années, c'est surtout par l'utilisation et l'interprétation des données de magnétisme sur des zones d'ouvertures océaniques.
Sur ces zones en effet, et depuis des millénaires, l'ouverture des fonds marins a mis à jour des magmas qui, au contact de l'eau, ont figé l'information "magnétisme" du moment.
L'examen du magnétisme sur ces zones équivaut ainsi à la lecture d'une bande magnétique sur laquelle apparaissent des alternances de signe correspondant aux changements de signe des pôles magnétiques (réf. [GΦ–6]).

La figure 1.5 contient un extrait de la zone SEAPERC que nous aurons le loisir d'étudier dans cette thèse. Cette carte montre des sillons alignés selon l'axe OX, et sensiblement parallèles. Si on observe une coupe perpendiculaire à ces sillons, on note l'alternance de signes dont il était question précédemment. (magnétisme : profil A–B). On notera que dans cet exemple, la variable gravimétrie suit les mêmes alternances, bien que plus irrégulièrement (gravimétrie : profil A – B).

Bien que cette structure soit, dans le plan X–Y, une réalité unique dont la taille réduite ne permet pas d'inférer une quelconque périodicité bi–dimensionnelle, l'examen de toutes les coupes "A–B" possibles lorsqu'on translate celles–ci parallèlement à l'axe OY permet d'obtenir ce que l'on peut considérer comme autant de réalisations d'une même "bande–magnétique" de cette dorsale.

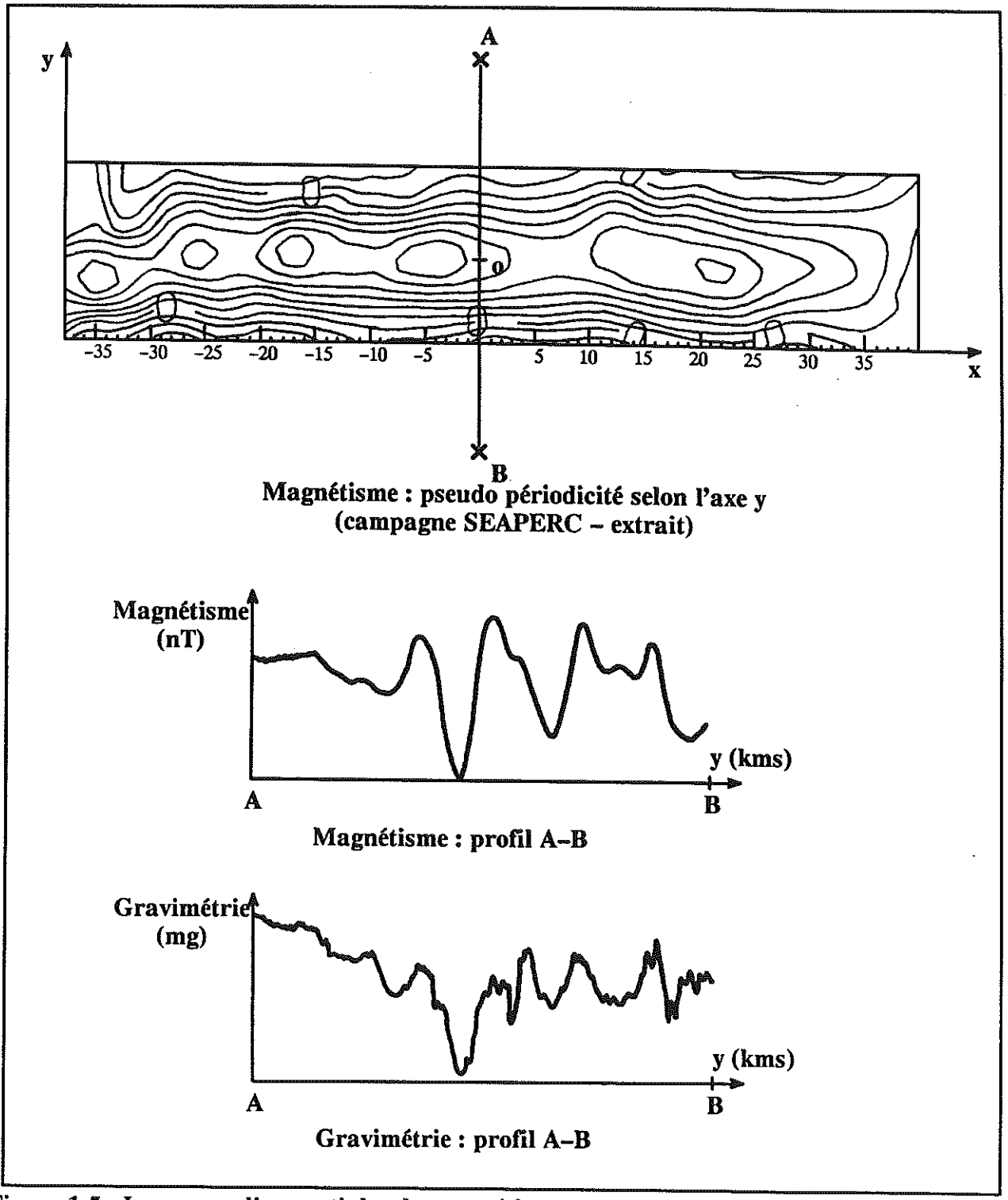


Figure 1.5 : Les anomalies spatiales du magnétisme et de la gravimétrie. Un exemple de structures spatiales "quasi" périodiques

Il peut s'avérer utile de considérer de telles zones comme un PTP, dont les caractéristiques sont:

- Espace de la VR : deux dimensions (le plan).
- Espace-période : l'axe OX.
- Période principale : variable selon la dorsale considérée.

1 .2 .3 Conclusions sur la section 1.2

Le catalogue qui vient d'être fait est loin d'être exhaustif. On aurait pu citer également les champs de séries pluviométriques, l'activité bactérienne, certaines perturbations en électronique ou encore la houle à la surface de l'eau. Cette tendance à "l'auto-répétition" est en fait très présente dans les VR, et si les conditions expérimentales le permettent (ie. nombre de périodes suffisant à l'échelle du champ d'observation de la VR), il peut s'avérer utile d'en tenir compte dans la modélisation.

Dans la suite, on travaillera surtout sur le magnétisme.

CHAPITRE 2

ANALYSE DE PTP TEMPORELS
(perturbations diurnes en magnétisme)

RÉSUMÉ

Ce chapitre contient les premiers exemples d'analyse de PTP : 3 séries temporelles de perturbations diurnes en magnétisme. Pour celles-ci, l'espace-période est confondu avec l'espace de régionalisation : le temps.

Selon la taille du champ informant, 3 schémas méthodologiques sont proposés pour la modélisation. Ils utilisent conjointement l'opérateur TFD et la fonction variogramme.

Ces méthodes permettront d'obtenir des ajustements très précis des covariances expérimentales.

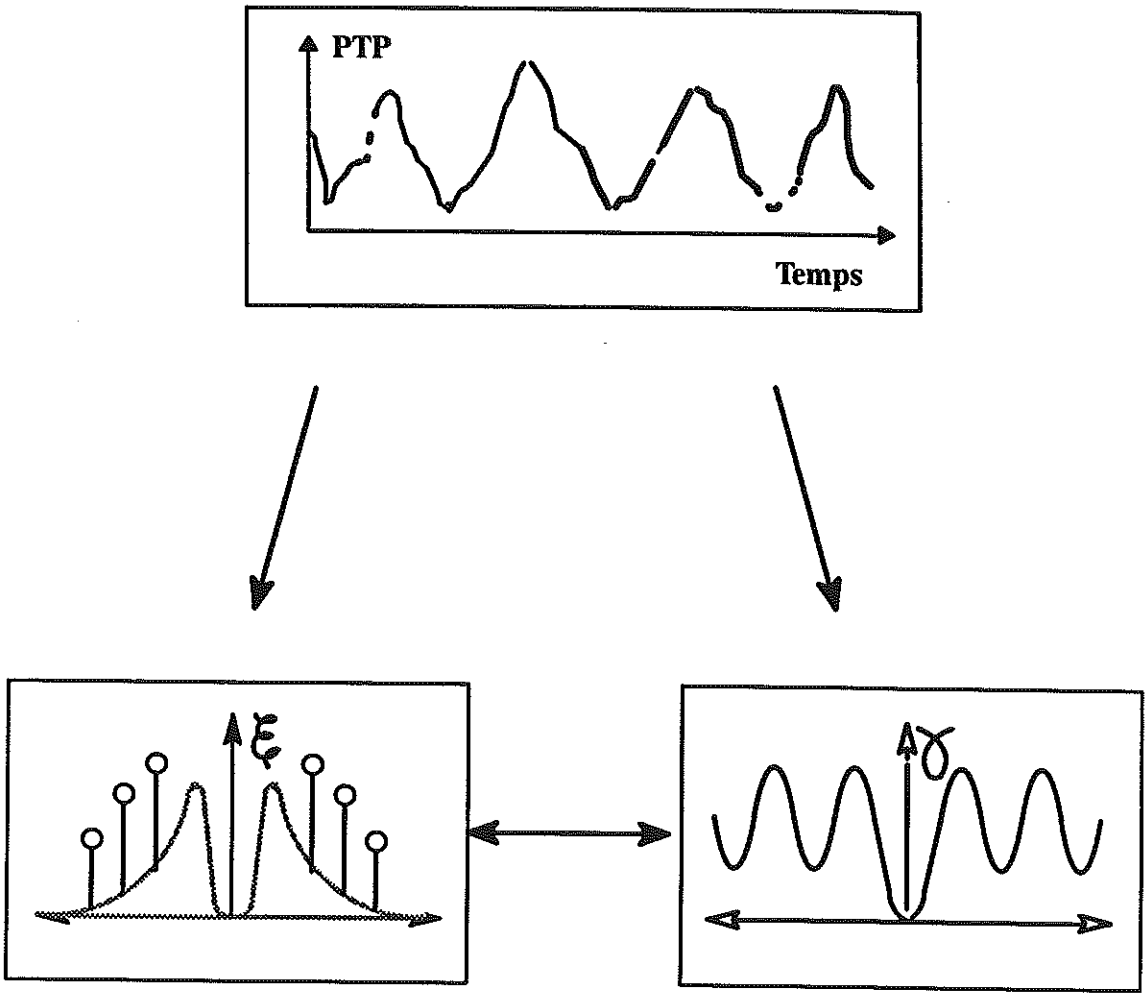
2.1 Préliminaires. page 29

2.2 Série temporelle à grande échelle d'analyse (KAKIOKA). page 32

2.3 Série temporelle à petite échelle d'analyse (HUANCAYO). page 48

2.4 Série temporelle à faibles oscillations (MOHNS). page 60

2.5 Conclusions du chapitre 2 . page 63


PTP
Temps
ξ
Δσ

Chapitre 2

ANALYSE DE PTP TEMPORELS (perturbations diurnes en magnétisme)

Dans ce chapitre, on s'attache à :

- Extraire du PTP les caractéristiques des fréquences qui émergent du spectre. La série de Fourier associée s'appellera alors "composante périodique" du PTP.

- Modéliser la covariance du complément que l'on appellera "composante non périodique".

Ces analyses s'appuient sur trois séries temporelles de mesures du magnétisme terrestre : Kakioka (section 2.2), Huancayo (section 2.3), et Mohns (section 2.4).

Bien que traitant du même phénomène physique (les perturbations diurnes du magnétisme terrestre, section 2.1), chacun des exemples possède sa spécificité puisque:

- La série Kakioka est connue sur 56 périodes principales, ce qui nous met dans des conditions expérimentales optimales pour appréhender la composante périodique.

- La série Huancayo n'est connue que sur 7 périodes principales, ce qui nous obligera à affiner nos techniques d'analyse.

- La série MOHNS n'est connue que sur 3 périodes et ne possède pas de période caractéristique.

Les analyses que nous effectuerons seront temporelles (calcul de variogrammes, et leur dichotomie) et spectrales (spectres d'amplitude et d'énergie).

Des configurations extrêmes d'échantillonnage (Kakioka et Huancayo), des fréquences qui n'émergent pas nécessairement dans le spectre (Mohns), autant de circonstances qui permettront de déterminer les conditions expérimentales optimales pour appréhender ce PTP temporel.

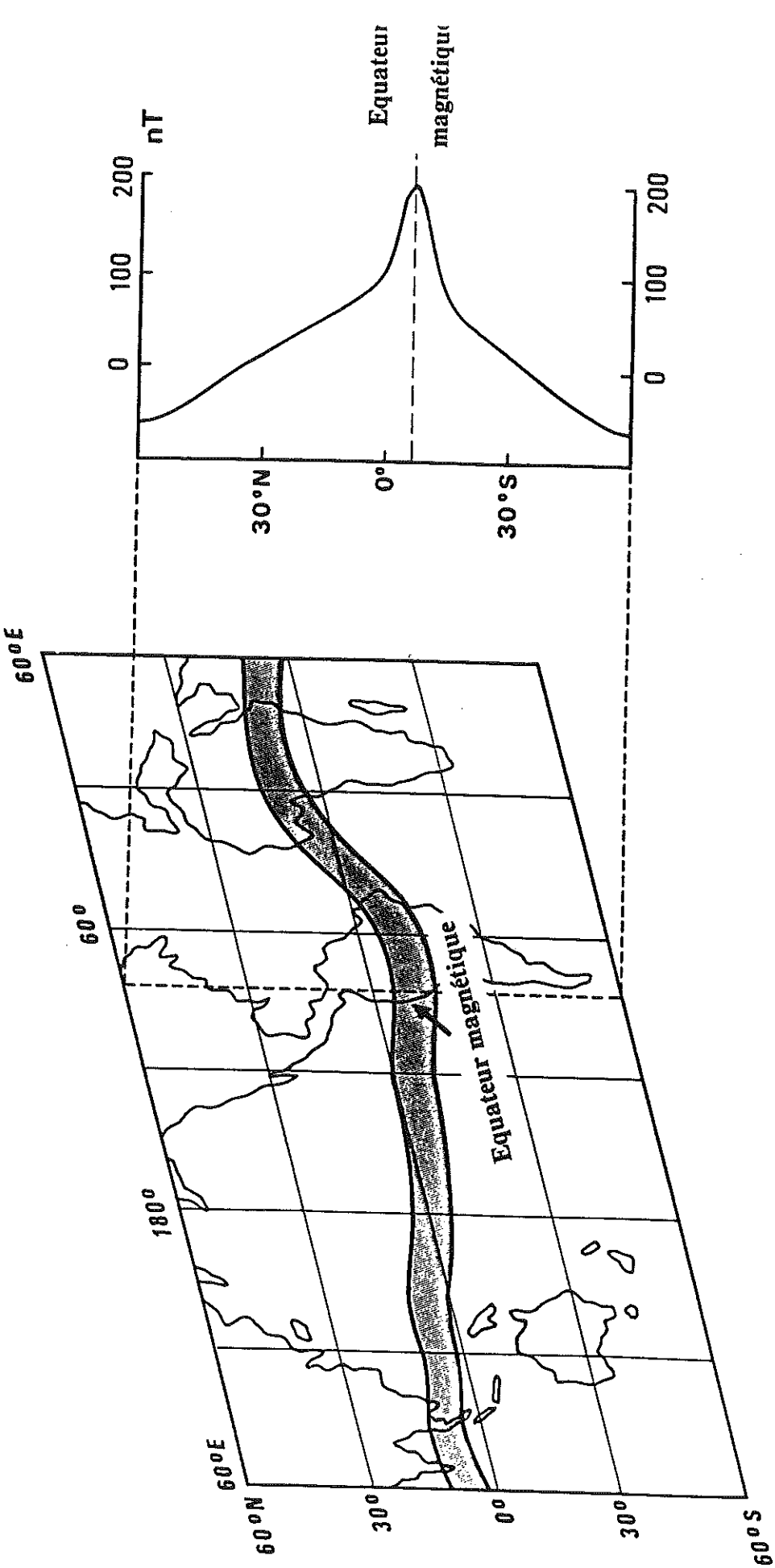


Figure 2.1 : Variations de la composante horizontale du champ magnétique en fonction de la latitude

2 .1 Préliminaires

2 .1 .1 Description géophysique des perturbations en magnétisme

Le champ géomagnétique ne dépend pas seulement de causes internes, mais aussi de causes externes, comme le vent solaire à l'origine des variations diurnes liées à la rotation de la terre (période 24 ^{h}) ou les orages magnétiques.

Ces variations (appelées "perturbations solaires") possèdent aussi des cycles annuels, avec un maximum en été (localement), un cycle semi-annuel de maximum près des équinoxes, et un cycle solaire lié à l'intensité du flux solaire (réf [GΦ-3]).

La rotation de la lune autour de la terre crée également des variations (période de 28 jours).

Les perturbations dépendent de la latitude du point d'observation et sont maximales à l'équateur magnétique où leurs variations crête à crête peuvent atteindre 200 nano-Tesla (nT) (réf [GΦ-7]). Sur la figure 2.1 sont tracés l'équateur magnétique ainsi que la composante horizontale de l'amplitude crête-à-crête des perturbations diurnes en fonction de la latitude. On notera l'importance croissante des perturbations lorsqu'on approche de l'équateur.

La variable qui sera explicitement analysée dans la suite sera celle du module du champ magnétique total.

2 .1 .2 Sur la TFD de covariances ou variogrammes

Dans les analyses qui suivent, nous serons parfois amené à effectuer des TFD (cf. annexe A) de covariances ou variogrammes expérimentaux pour évaluer les énergies en jeu dans le signal et effectuer des ajustements.

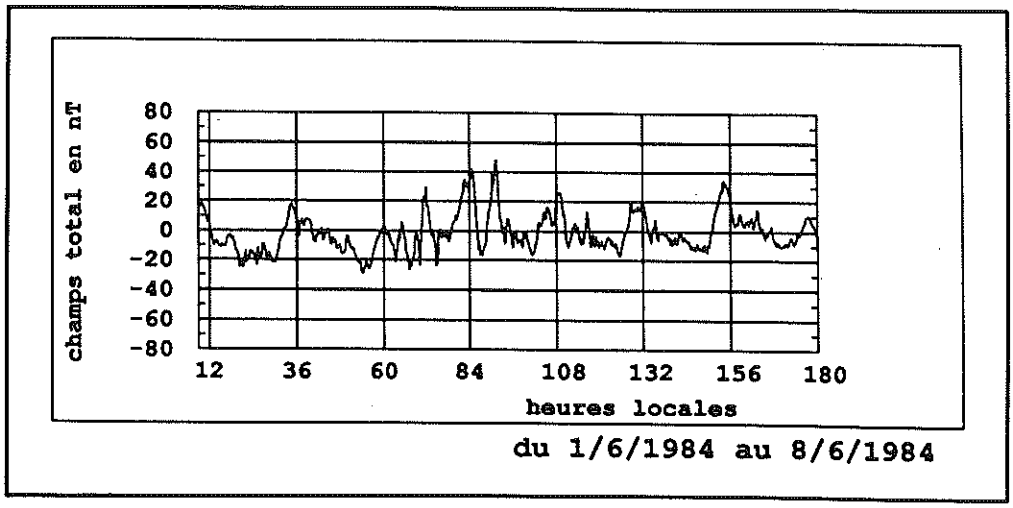
Il est bon de rappeler que les covariances expérimentales que nous manipulons ne sont pas des fonctions définies positives, ceci essentiellement parce que leur calcul donné par :

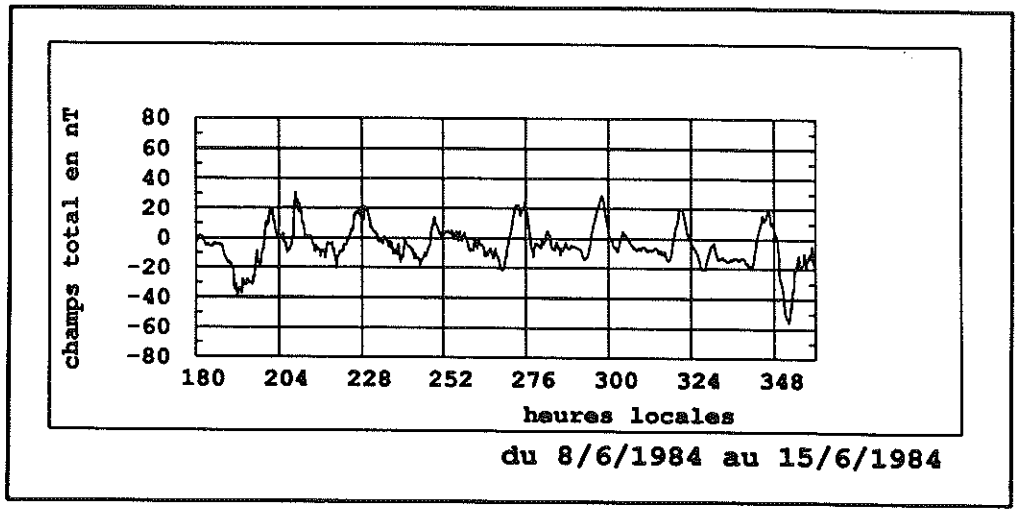
$$C^{*}(h) = \frac{1}{\mu(h)} \int z(x+h)z(x)dx$$

fait intervenir au dénominateur la mesure du domaine d'intégration dont la Transformée de Fourier ($T\mathcal{F}$) détruit la positivité de celle du numérateur. C'est seulement pour un domaine s'étendant à l'infini que la positivité peut apparaître et que le théorème de Bochner peut donner par $T\mathcal{F}$ une mesure positive sommable.

Il en résulte que lorsque nous calculerons de telles TFD, des énergies négatives peuvent apparaître. Si tel est le cas, nous renoncerons à l'interprétation des spectres associés (sauf bien entendu si ces quantités négatives sont négligeables).

Il en est de même de la TFD de variogrammes expérimentaux qui, même manifestement stationnaires, peuvent engendrer des énergies positives, alors qu'une fonction variogramme (théorique et non expérimentale) stationnaire n'engendre que des énergies négatives.





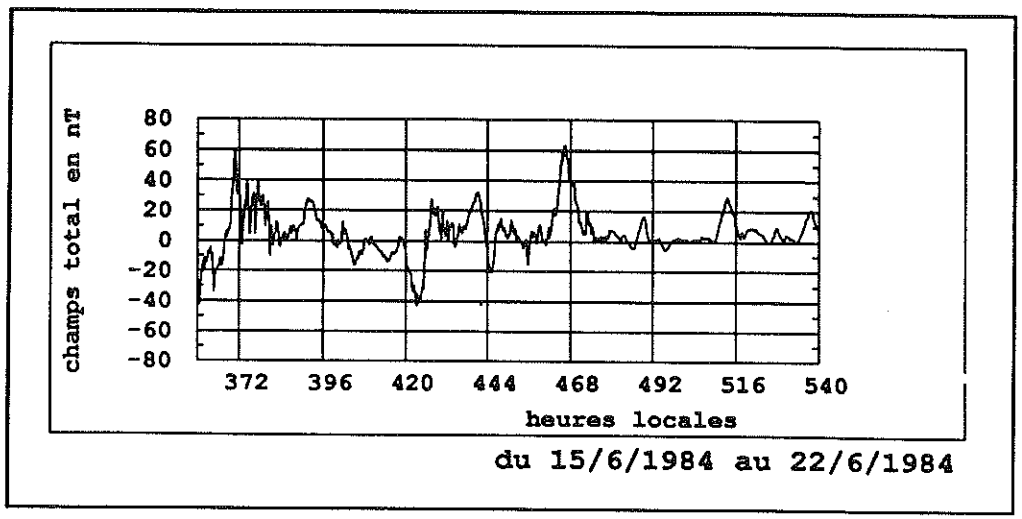
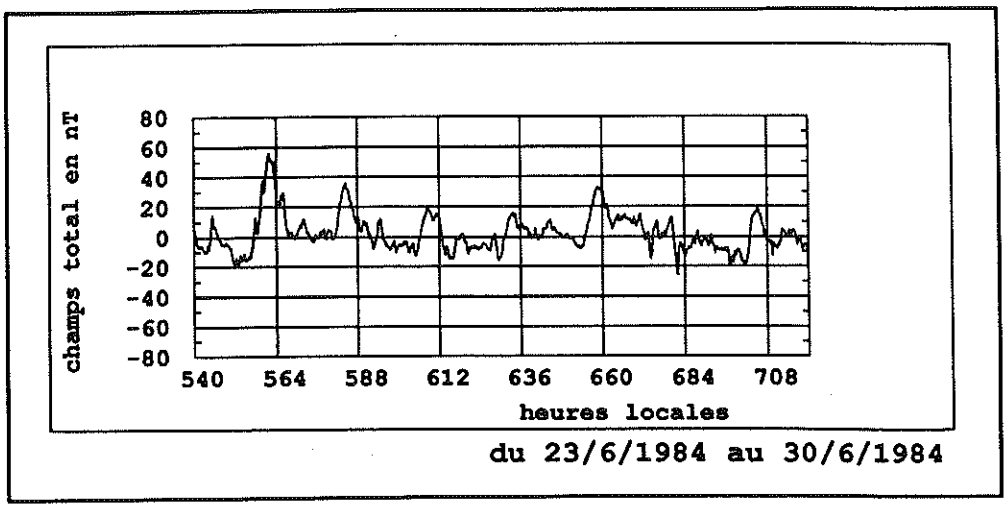
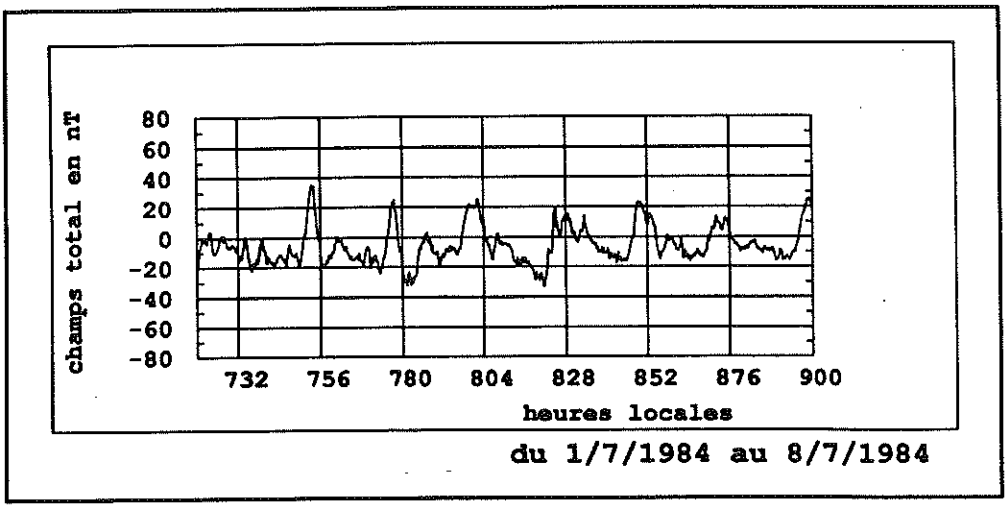


Figure 2.2 : Evolutions dans le temps du champ magnétique total à l'observatoire fixe de Kakioka (Japon)





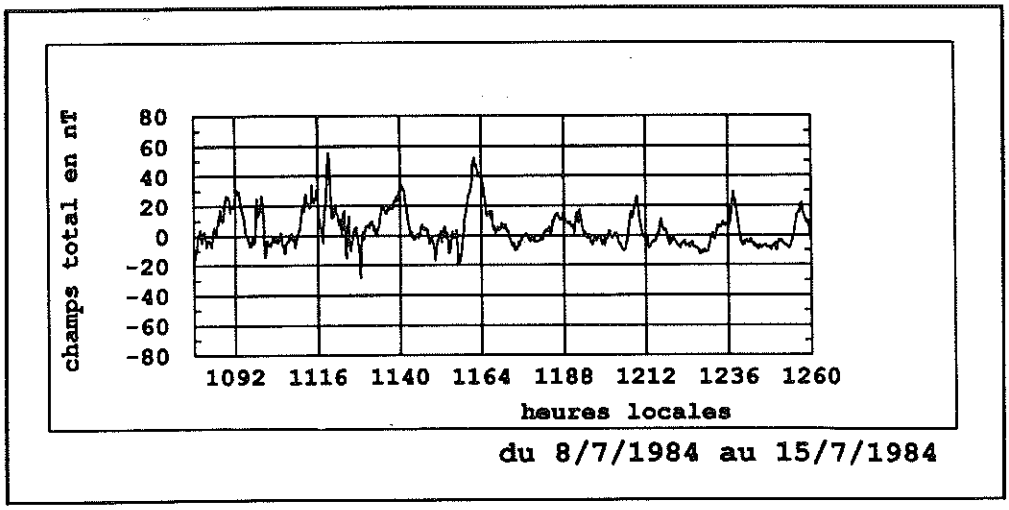


Figure 2.2 (suite et fin)

2 .2 Série temporelle à grande échelle d'analyse : Kakioka

2 .2 .1 Présentation des données

Ces données ont été mesurées à l'observatoire fixe de Kakioka, au Japon ($\simeq$ 40 º N, 140 º E) et nous ont été aimablement fournies par l'intermédiaire de Ph. Huchon (réf [I-2]).

Elles ont été enregistrées toutes les minutes du 1/6/1984 au 30/7/1984 ce qui représente près de 85 000 points de mesure. La mesure est celle, en nT, de l'amplitude du vecteur champ magnétique total, translatée de telle sorte que la moyenne des échantillons soit nulle. Nous appellerons D(t) cette variable régionalisée dans le temps.

2 .2 .2 Courbe expérimentale de D(t)

Les courbes de la figure 2.2 montrent l'évolution de D(.) au fil d'un temps exprimé en heures locales. On notera que ce signal est très erratique et que la pseudo-périodicité est loin d'être immédiate, surtout pour les 7 premiers jours de mesure. On note cependant que les multiples de 11 heures représentent en moyenne le moment où D(.) atteint ses maxima.

Le temps local est ici déphasé de 1 heure par rapport au temps solaire, ce qui fait que le soleil est au zénith à 11 heures locales. C'est à ce moment de la journée que l'interaction du flux solaire et du champ magnétique est maximale.

On notera que les zones de nuit sont relativement moins actives, mais demeurent bruitées. Le domaine de variation de D(.) est de 60 nT à - 40 nT, soit une amplitude crête à crête de 100 nT.

Ceci peut paraître excessif en regard de la latitude (40º N) du point de mesure, si l'on se remémore les variations de ces amplitudes en fonction de la latitude (figure 2.1). Mais sur la figure 2.1, c'est la composante horizontale du champ qui était mesurée et visualisée en fonction de la latitude.

En résumé, un examen visuel de ces courbes ne révèle pas grand chose quant au caractère périodique d'une VR qu'il est encore prématuré d'appeler PTP.

2 .2 .3 Périodogramme de D(.)

Puisqu'on suspecte la présence d'oscillations dans ce processus, nous lui avons appliqué une analyse de Fourier par TFD.

On rappelle (cf. annexe A) que cette opération permet de décomposer la fonction en série de Fourier, selon la formule :

$$D(t_\alpha, D_{TFD}) = \sum_i [\ A_i(D_{TFD}) \cos(\omega_i t_\alpha) + B_i(D_{TFD}) \sin(\omega_i t_\alpha)\ ]\ , \qquad (2.1)$$

avec $\omega_i = k\ \omega_\beta$, $\omega_\beta = \dfrac{2\pi}{D_{T.F.d.}}$, et pour D_{TFD} la longueur du champ analysé.

On rappelle également que la résolution fréquentielle du résultat de l'analyse dépend de la taille du champ et du pas d'échantillonnage p_{TFD} de la fonction .

Fréquence ω_i	Périodes (minutes) T_i	Amplitude cosinus A_i	Amplitude sinus B_i	$\sqrt{A_i^2 + B_i^2}$	Phase (π radians) ϕ	
0.0000740	84960.00	-1.15	-1.14	1.62	1.15	
0.0001479	42480.00	4.28	-2.74	5.08	1.18	→ $708^h = 29.5\,j$
0.0002279	28320.00	0.96	4.93	5.02	0.44	→ $708^h = 29.5\,j$
0.0002958	21240.00	0.45	0.95	1.05	0.36	
0.0003698	16992.00	-0.51	-1.26	1.36	1.38	
0.0004437	14160.00	-1.39	-0.38	1.44	1.09	
0.0005177	12137.14	0.44	1.67	1.72	0.42	
0.0005916	10620.00	-0.93	-1.75	1.98	1.34	
0.0006656	9440.00	-2.18	-0.35	2.21	1.05	
0.0008135	7723.64	1.81	-0.49	1.88	-0.08	
0.0008875	7080.00	-0.29	-1.46	1.49	1.44	
0.0009614	6535.38	-0.75	1.30	1.51	0.67	
0.0011093	5664.00	2.22	-1.73	2.81	-0.21	
0.0011833	5310.00	-0.52	0.91	1.05	0.66	
0.0012572	4997.65	-1.23	-0.68	1.40	1.16	
0.0013312	4720.00	0.94	-0.94	1.33	-0.25	
0.0014051	4471.58	1.98	-1.16	2.29	-0.17	
0.0014791	4248.00	0.21	1.97	1.98	0.47	
0.0015530	4045.71	-1.71	1.12	2.04	0.82	
0.0016270	3861.82	1.87	0.04	1.87	0.01	
0.0017749	3540.00	-1.91	1.09	2.20	0.84	
0.0019228	3267.69	-0.16	-1.32	1.33	1.46	
0.0020707	3034.29	0.48	1.16	1.25	0.37	
0.0021447	2929.66	-0.99	0.18	1.01	0.94	
0.0022186	2832.00	0.13	-1.41	1.42	-0.47	
0.0022926	2740.65	-0.41	2.16	2.20	0.56	
0.0024405	2574.55	0.20	-1.50	1.51	-0.46	
0.0025145	2498.82	1.14	-1.34	1.76	-0.28	

Tableau 2.1 : Amplitudes aux fréquences de D(.) (série temporelle Kakioka)

Fréquence ω_i	Périodes (minutes) T_i	Amplitude cosinus A_i	Amplitude sinus B_i	$\sqrt{A_i^2 + B_i^2}$	Phase (π radians) ϕ	
0.0026624	2360.00	1.20	–0.32	1.24	–0.08	
0.0031800	1975.81	2.84	–0.09	2.84	–0.01	
0.0034019	1846.96	1.20	–0.84	1.46	–0.19	
0.0038456	1633.85	0.07	1.00	1.00	0.48	
0.0039196	1603.02	–1.09	–0.51	1.20	1.14	
0.0039935	1573.33	0.79	1.55	1.74	0.35	
0.0041415	1517.14	1.09	–0.83	1.37	–0.21	
0.0043633	1440.00	6.38	4.98	8.10	0.21	→ 24^h
0.0045112	1392.79	–2.11	–0.57	2.18	1.08	
0.0045852	1370.32	–1.42	–0.11	1.42	1.02	
0.0047331	1327.50	0.94	0.76	1.21	0.22	
0.0051029	1231.30	1.13	0.03	1.13	0.01	
0.0054726	1148.11	1.28	0.59	1.41	0.14	
0.0066559	944.00	0.20	1.38	1.39	0.45	
0.0071736	875.88	1.07	–0.05	1.07	–0.01	
0.0074434	832.94	0.96	0.51	1.09	0.16	
0.0087266	720.00	5.55	1.78	5.83	0.10	→ 12^h
0.0090964	690.73	–1.17	0.25	1.19	0.93	
0.0091704	685.16	–1.32	0.49	1.41	0.89	
0.0099839	629.33	1.27	0.48	1.36	0.12	
0.0105755	594.13	–0.80	0.87	1.18	0.74	
0.0113890	551.69	0.34	–1.17	1.21	–0.41	
0.0130900	480.00	3.10	2.89	4.24	0.24	→ 8^h
0.0142732	440.21	–0.49	1.09	1.19	0.64	
0.0153826	408.46	–0.05	1.19	1.19	0.51	
0.0218166	288.00	0.98	–0.84	1.29	–0.23	

Tableau 2.1 (suite et fin)

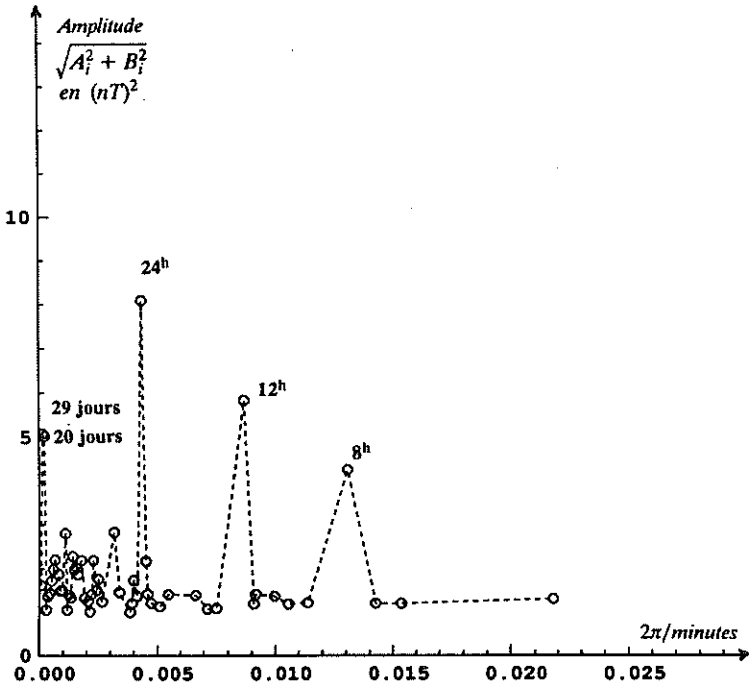


Figure 2.3 : Spectre d'amplitude de D(.)

Périodes (minutes) T_i	Amplitude en nT	
29.5 jours	5.08 nT	Origine : rotation de lune autour de la terre
19.6 jours	5.02 nT	
24 heures	8.10 nT	Origine : rotation de la terre sur elle même
12 heures	5.83 nT	
8 heures	4.24 nT	

Tableau 2.2 : Les fréquences qui émergent du spectre d'amplitude de D(.)

Pour cette première analyse, nous avons adopté les paramètres suivants :

- D_{TFD} = 59 jours
- P_{TFD} = 10 minutes

La figure 2.3 représente le spectre d'amplitude obtenu par la formule:

$$\text{amplitude (i)} = \sqrt{A_i(D_{TFD})^2 + B_i(D_{TFD})^2} \qquad A_i(.) \text{ et } B_i(.) \text{ issus de 2.1} \tag{2.2}$$

et les tableaux 2.1 fournissent le détail des chiffres. N'y figurent que les fréquences possédant une amplitude supérieure à 1 nT.

On notera que le tableau 2.1 contient également les phases ϕ_i associées aux fréquences..
Pour une fréquence ω_i donnée, on a en effet,

$$A_i\cos(\omega_i t) + B_i\sin(\omega_i t)$$

$$= \sqrt{A_i^2 + B_i^2}\left(\frac{A_i}{\sqrt{A_i^2 + B_i^2}}\cos\omega_i t + \frac{B_i}{\sqrt{A_i^2 + B_i^2}}\sin\omega_i t\right)$$

$$= \sqrt{A_i^2 + B_i^2}\ \cos(\omega_i t - \phi_i)$$

avec ϕ_i tel que $\cos(\omega_i\phi_i) = \frac{A_i}{\sqrt{A_i^2 + B_i^2}}$ *et* $\sin(\omega_i\phi_i) = \frac{B_i}{\sqrt{A_i^2 + B_i^2}}$

Nous n'utiliserons pas ces spectres de phase dans la suite bien qu'ils constituent une information essentielle et un moyen de quantifier ce que l'on perd du processus lorsqu'on choisit de le caractériser par son spectre d'énergie ou sa covariance.

L'examen du spectre d'amplitude de D(.) montre l'émergence de 5 périodes (tableau 2.2), révélatrices de deux phénomènes physiques d'échelles différentes :

- les perturbations lunaires : période principale : 28 jours
- les perturbations diurnes : période principale : 24 heures

Il convient de les étudier séparément :

Les perturbations lunaires :

Le champ d'analyse (59 jours) contient un peu plus de 2 périodes seulement de ce phénomène, ce qui est trop peu pour que l'analyse par TFD s'avère précise. De plus 59 jours n'est pas un multiple de 28 jours. Or, le peigne fréquentiel associé aux paramètres ($D_{TF.d.}$, P_{TFD}) est composé des périodes T_i données par (rappel) :

$$T_i = \frac{D_{TFD}}{i} \qquad \text{avec i variant de } \emptyset \text{ à } \frac{D_{TFD}}{p_{TFD}}\ . \tag{2.3}$$

Il en résulte que "28 jours" ne fait pas partie des "périodes possibles" de l'analyseur. L'énergie associée à la lune se répartit alors sur les fréquences voisines: c'est le phénomène de diffusion bien connu en Théorie du Signal (cf. annexe A–3).

Comme nous sommes essentiellement intéressé par les perturbations diurnes, nous n'avons pas renouvelé l'expérience avec un peigne fréquentiel adapté.

Les perturbations diurnes :

La situation est toute autre pour les perturbations diurnes puisque le champ d'analyse contient 59 périodes principales. On peut donc penser que la période 24 heures est correctement quantifiée par cette TFD .

Il en est de même, bien entendu, des périodes 12 heures et 8 heures, dont on ne peut dire pour l'instant, si elles sont liées à l'influence de la terre ou de la lune. Pour trancher, il faudra réduire considérablement le champ d'analyse jusqu'à pouvoir négliger l'influence de la lune.
Si ces deux fréquences persistent, nous pourrons alors en déduire qu'elles appartiennent à l'influence de la terre.

En effet, on pourrait penser que ces périodes 12 h et 8 h sont issues du phénomène de diffusion précédemment évoqué
Le problème ici est que si un champ long de près de 2 x 28 jours est insuffisant pour spécifier convenablement l'influence de la lune, c'est cependant trop grand pour

pouvoir négliger complètement celle-ci. Le doute se lèvera de lui-même lorsque nous effectuerons des TFD sur des champs de quelques jours seulement.

2 .2 .4 Variogramme expérimental de D(.)

L'abondance de l'information va permettre d'examiner son variogramme (noté $\gamma^*_D(.)$) à différentes échelles. Cette fonction a été calculée par pas de 100 ou 10 minutes selon l'échelle d'analyse.

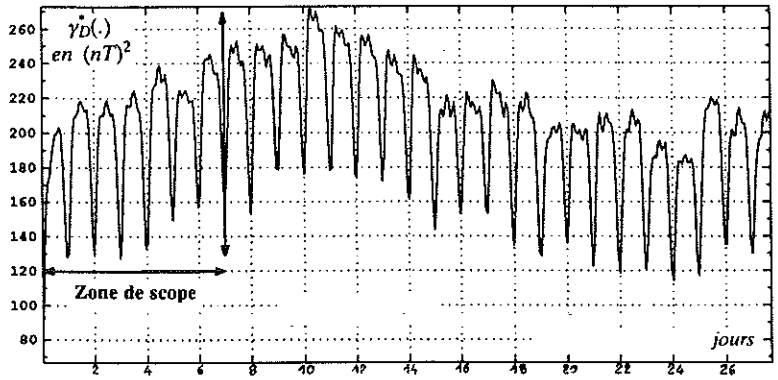


Figure 2.4 : Variogramme expérimental des perturbations (25 jours)

La figure 2.4 révèle que $\gamma^*_D(.)$ possède une oscillation principale de période supérieure à 25 jours et sur laquelle reposent des oscillations très régulières de période égale à 24 heures. Bien que le champ total ne contienne que 2 périodes lunaires, ce phénomène est cependant correctement représenté par $\gamma^*_D(.)$.

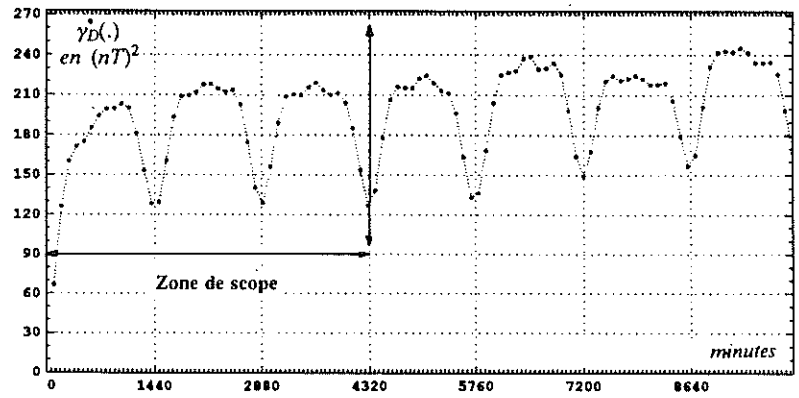


Figure 2.5 : Variogramme expérimental des perturbations (7 jours)

L'examen de $\gamma^*_D(.)$ à l'échelle de 7 jours permet de mieux appréhender l'allure des oscillations de 24 heures. L'analyse de cette courbe révèle en effet, 3 zones importantes dans ce variogramme:

- Pour h variant de $\emptyset$ à 1 jour, on note la "montée" de $\gamma_D(h)$ et une première oscillation.

- Pour h variant sur les jours 2 et 3, on note deux oscillations identiques et reposant sur un palier égal à 120 $(nT)^2$. Nous sommes ici dans une zone de "transition" où les oscillations terrestres sont parfaitement structurées, mais où la lune n'a pu encore manifester son influence.
- Pour h variant de 3 à 7 jours, on note la naissance de l'influence lunaire, puisque le palier des oscillations terrestres commence à croître.

L'existence de ce "no men's land" des influences terrestres et lunaires (jours 2 et 3) est une circonstance tout à fait remarquable qui va permettre d'effectuer un ajustement très précis de cette fonction expérimentale.

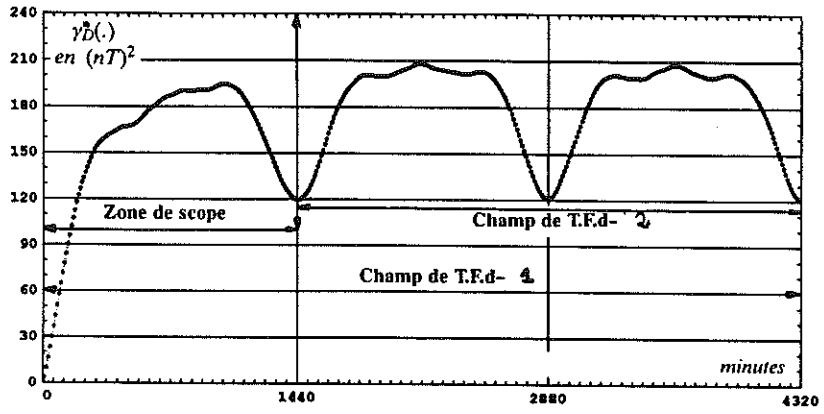


Figure 2.6 : Variogramme expérimental des perturbations (3 jours)

L'examen de $\gamma^*_D(.)$ à l'échelle de 3 jours (figure 2.6) confirme nos remarques précédentes, à savoir l'existence d'un palier "local" (pour h variant de 1 à 3 jours) égal à 120 $(nT)^2$ et sur lequel reposent deux ondulations similaires.

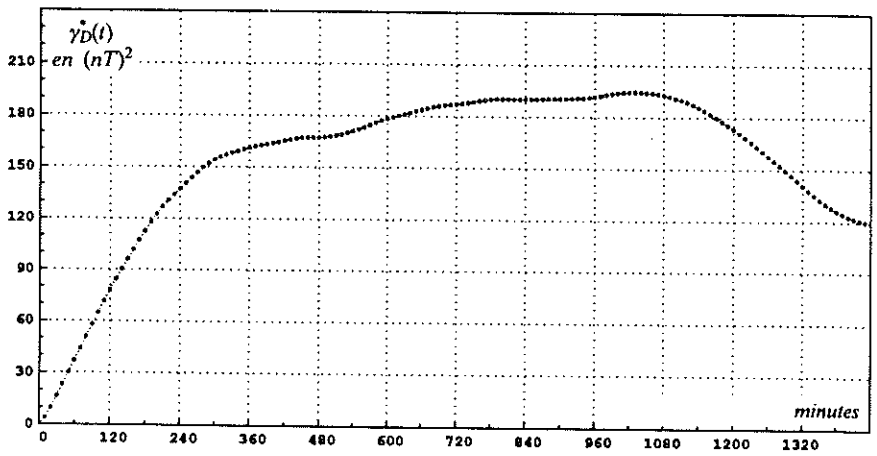


Figure 2.7 : Variogramme expérimental des perturbations (1 jours)

L'examen de $\gamma^*_D(.)$ à l'échelle de 24 heures (figure 2.7) montre un variogramme ayant un comportement linéaire à l'origine.

En conclusion, l'examen à plusieurs échelles de $\gamma^*_D(.)$ révèle l'existence de 3 phénomènes très structurés au sein du même processus D(.) :

- un phénomène non périodique d'échelle inférieure à 24 heures qui, s'il s'avère modélisable à l'aide d'une structure à palier (notée $\sigma_D(.)$), possèdera un palier proche de 120 $(nT)^2$.
- un phénomène périodique de période principale 24 h, qui contient également d'autres périodes caractéristiques (12 h et 8 h),
- un phénomène périodique de période principale supérieure à 25 jours, et dont on ignore s'il contient ou non d'autres fréquences caractéristiques.

Parmi ces trois phénomènes, il en existe deux qui ont des zones d'influences qui ne s'intersectent pas puisqu'avant que la période 28 jours ne se révèle, celle de 24 h s'est déjà réalisée trois fois (le "no men's land" déjà évoqué).

2 .2 .5 Spectre d'énergie de D(.)

La longueur du champ d'analyse de γ^*_D est de 3 jours. Comme le variogramme est une fonction paire, une telle TFD ne doit donner une décomposition que sur des fonctions cosinus. Encore faut-il fournir à l'analyseur le moyen de constater cette parité. C'est pourquoi le signal d'entrée de la TFD est $\gamma_D(h)$ pour h variant de -3 jours à + 3 jours. En conséquence, D_{TFD} est ici de 6 jours et P_{TFD}, vaut 10 minutes.

Le tableau 2.3 montre des coefficients tous négatifs (sauf bien entendu la constante de Fourier). Il en résulte qu'avec ces paramètres D_{TFD} et p_{TFD}, le variogramme expérimental γ^*_D peut être considéré comme la version discrète d'une fonction variogramme stationnaire, identifiable à la série de Fourier :

$$\gamma_D(h) = \sum_i \xi(\omega_i)(1 - \cos(\omega_i h)) \tag{2.4}$$

L'inférence de la covariance $C_D(.)$ qui en résulte est immédiate, et l'on a :

$$C_D(h) = \sum_i \xi(\omega_i)\cos(\omega_i h), \tag{2.5}$$

avec

$$C_D(\emptyset) = \sum_i \xi(\omega_i) \;=\; \text{constante de Fourier} \;=\; \text{VAR (D)} \tag{2.6}$$

La figure 2.8 contient le spectre d'énergie $\xi(\omega_i)$ associé à $C_D(.)$. Seules les énergies supérieures à 1 nT y figurent. Ce spectre est tout à fait remarquable, tant il caractérise le PTP : des périodes 24 h, 12 h et 8 h qui émergent d'un spectre qui de plus décroît rapidement vers $\emptyset$. De ce fait, on pressent dans ce spectre, la présence de deux caractéristiques structurales :

- la composante périodique caractérisée par les périodes 24 h, 12 h et 8 h.
- la composante non périodique qui constitue en quelque sorte l'information portée par la composante périodique.

On se reportera au chapitre introductif où nous parlions de "modulation à porteuse sinusoïdale" pour interpréter l'image suivante : la perturbation diurne est un signal modulé par une porteuse sinusoïdale de caractéristiques fréquentielles 24 h, 12 h et 8 h, et dont l'information portée possède un spectre à bande étroite.

Fréquence ω_i	Périodes (minutes) T_i	-Energie $\xi = TFD(\gamma_D^*)$	
0.0000000	0.00	173.28	→ *Constante de Fourier*
0.0007272	8640.00	-17.57	→ *Fréquences émergentes*
0.0014544	4320.00	-13.94	→ *Fréquences émergentes*
0.0021817	2880.00	-9.50	
0.0029089	2160.00	-6.93	
0.0036361	1728.00	5.86	
0.0043633	1440.00	-37.77	→ *Fréquences émergentes*
0.0050905	1234.29	-3.65	
0.0058178	1080.00	-2.46	
0.0065450	960.00	-2.37	
0.0072722	864.00	-2.29	
0.0079994	785.45	-1.75	
0.0087266	720.00	-19.91	→ *Fréquences émergentes*
0.0095539	664.62	-2.52	
0.0101811	617.14	-1.95	
0.0109083	576.00	-2.63	
0.0116355	540.00	-2.18	
0.0126355	508.24	-1.16	
0.0130900	480.00	-10.00	→ *Fréquences émergentes*
0.0138172	454.74	-1.19	
0.0145444	432.00	-1.37	
0.0152716	411.43	-1.76	
0.0159989	392.73	-1.62	
0.0167261	375.65	-1.23	
0.0174533	360.00	-1.45	
0.0203622	308.57	-1.14	
0.0218166	288.00	-1.28	

Tableau 2.3 : Energies aux fréquences associées aux perturbations

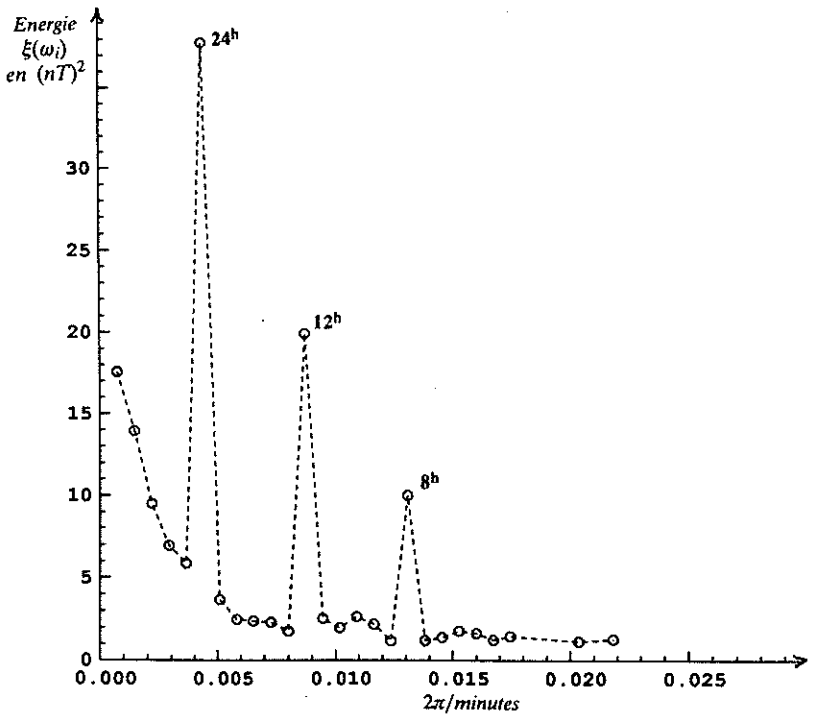


Figure 2.8 : Spectre d'énergie associé aux perturbations

2 .2 .6 Anatomie d'un ajustement

Dans cette section, nous allons pratiquer l'inférence d'un modèle de variogramme à partir de l'expérimental $\gamma^*_D(.)$, par une utilisation conjointe de la TFD et des techniques géostatistiques usuelles d'ajustement.
Comme nos préoccupations portent sur l'influence terrestre (et non celle de la lune), la zone sur laquelle s'effectuera cet ajustement sera telle que :

- les perturbations diurnes soient présentes,
- les perturbations lunaires soient négligeables.

Or, il ressort de l'analyse précédente qu'à l'échelle de h $\in$ [0, 3 jours], $\gamma^*_D(h)$ présente pour les jours 2 et 3, une périodicité stricte par 2 ondulations reposant sur le même palier (figure 2.6) Nous en avions déduit qu'à cette échelle, la lune n'avait pas encore manifesté son influence . Notre échelle d'analyse sera donc cette zone [0, 3 jours].

Cette inférence procède de 4 étapes.

Etape 1, TFD de γ^*_D sur les jours 2 et 3 :

Il s'agit de savoir qu'elles sont les fréquences et énergies qui composent les ondulations des jours 2 et 3. Si on utilisait, dans ce but, la TFD de γ^*_D (zone TFD-1 sur la figure 2.6) sur les trois premiers jours, on intègrerait dans cette analyse une partie de l'énergie qui appartient à σ^*_D (la composante non périodique de γ^*_D).

En effet, l'examen du spectre de la figure 2.8 montre que si les fréquences 24 h, 12 h et 8 h émergent bien du spectre, ce ne sont pas les valeurs 37.77, 19.91 ou 8.79 qu'il faut prendre pour modéliser les oscillations des jours 2 et 3, mais plutôt ces valeurs réduites de la valeur, à chaque f_i *(i = 1, 3) de la contribution de* σ^*_D *, soit approximativement 5, 3 et 1* $(nT)^2$ *pour 24 h, 12 h et 8 h.*

Or, notre objectif est d'effectuer une dichotomie en une composante périodique et une structure σ^*_D. Pour détecter le plus précisément possible les caractéristiques spectrales de cette composante périodique, il s'avère nécessaire d'effectuer cette TFD sur une zone où σ^*_D s'est

accompli, et ne perturbe plus la composante périodique du variogramme. Comme la nature offre ici une telle zone (les jours 2 et 3), la TFD du tableau 2.4 (qui constitue le premier pas dans notre ajustement) a été faite sur $\gamma^*_D(h)$, pour h variant sur les jours 2 et 3. (zone TFD-2 sur la figure 2.6).

Fréquence ω_i	Périodes (minutes) T_i	-Energie ξ = TFD(γ^*_D)	
0.0000000	0.00	182.23	→ *Constante de Fourier*
0.0010908	5760.00	-1.58	
0.0021817	2880.00	-1.13	
0.0032725	1920.00	-2.52	
0.0043633	1440.00	-33.63	→ 24^h
0.0087266	720.00	-17.86	→ 12^h
0.0130900	480.00	-8.79	→ 8^h

Tableau 2.4 : Analyse de la composante périodique du variogramme

On y constate que toutes les énergies sont négatives (c'est un variogramme qui est analysé) de sorte que cette fonction expérimentale peut s'identifier à la version échantillonnée d'un modèle périodique. Si on effectue une sommation des valeurs absolues de ces énergies, on obtient une valeur de 65.51 qui représente l'énergie (ou la variance) de la composante périodique pure de $\gamma^*_D(.)$. Si on retire cette quantité de la constante de Fourier (égale ici à 182,23), on obtient une valeur de 116.72. Cette valeur représente tout simplement une estimation de la variance de la composante non périodique du PTP.

En ce qui concerne les énergies que l'on peut associer aux fréquences qui émergent, elles sont:

- Pour $T_1 = 24^h$, $\xi(\omega_1) = 33.63(nT)^2$
- Pour $T_2 = 12^h$, $\xi(\omega_2) = 17.86(nT)^2$
- Pour $T_3 = 8^h$, $\xi(\omega_3) = 8.79(nT)^2$

Si on compare ces énergies à celles déduites de la TFD de γ^*_D sur 3 jours (zone TFD-1 sur la figure 2.6, et tableau 2.3 pour les chiffres), on constate que ces valeurs sont (en énergie équivalente) toutes inférieures. Cette TFD sur les jours 2 et 3 seulement a donc permis de dissocier l'énergie de σ^*_D , de celle de la composante périodique pure.

Maintenant en possession des caractéristiques spectrales exactes de la composante périodique de D(.), il est possible d'en déduire un modèle périodique de variogramme en appliquant la formule 2.4 avec comme paramètres:

- $\xi_1 = 33.63$ et $\omega_1 = 2\pi$ / 24 heures
- $\xi_2 = 17.86$ et $\omega_2 = 2\pi$ / 12 heures
- $\xi_3 = 8.79$ et $\omega_3 = 2\pi$ / 8 heures

Nous appellerons " $\gamma_{pério}$ (h) " cette structure . La figure 2.9 représente les évolutions de la fonction $\gamma_{pério}$ ainsi définie. Ces évolutions reproduisent bien entendu (à une constante près) ce qui se passait dans $\gamma^*_D(h)$, pour h variant sur les jours 2 et 3.

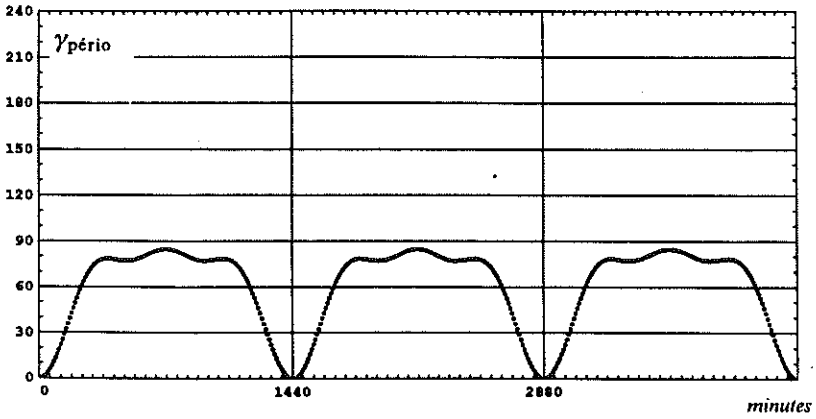


Figure 2.9 : Composante périodique du variogramme

Etape 2, extraire σ_D^* :

Puisque $\gamma_{pério}(.)$ est modélisé, il reste à extraire $\sigma_D^*(.)$, en effectuant l'opération suivante :

$$\sigma^*_D = \gamma^*_D(h) - \gamma_{pério}(h) \qquad (2.7)$$

pour h variant de $\emptyset$ à 3 jours.

Cette fonction, qui est déjà un modèle puisque $\gamma_{pério}$ en est un, est représentée sur la figure 2.10.

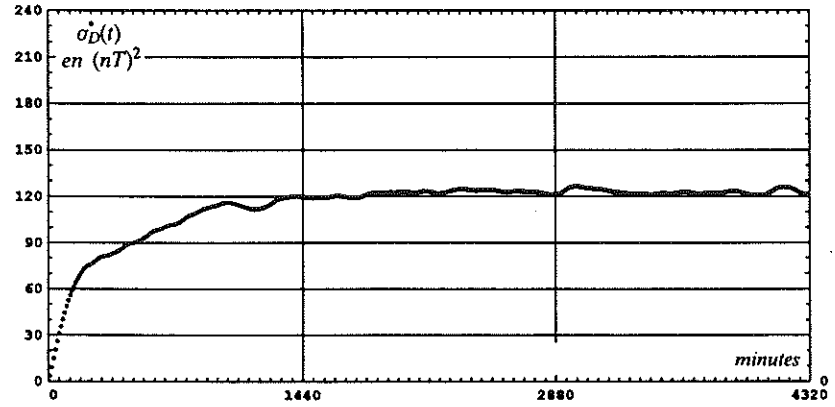


Figure 2.10 : Composante non périodique du variogramme

Elle possède les évolutions d'un phénomène à palier. A première vue, ce palier semble atteint aux alentours de 24 h.

Etape 3, modéliser σ_D^* :

Il reste à ajuster σ_D^*, manuellement, comme pratiqué en géostatistique. Pour cette modélisation, les 3 structures suivantes ont été utilisées :

- Modèle sphérique de portée a_1 = 200 minutes et de palier 60 $(nT)^2$
- Modèle sphérique de portée a_2 = 1100 minutes et de palier 43 $(nT)^2$
- Modèle gaussien de portée pratique a_3 = 1732 minutes (paramètre égal à 1000) et de palier égal à 20 $(nT)^2$.

Ceci donne une valeur de 123 $(nT)^2$ par le palier de σ_D.

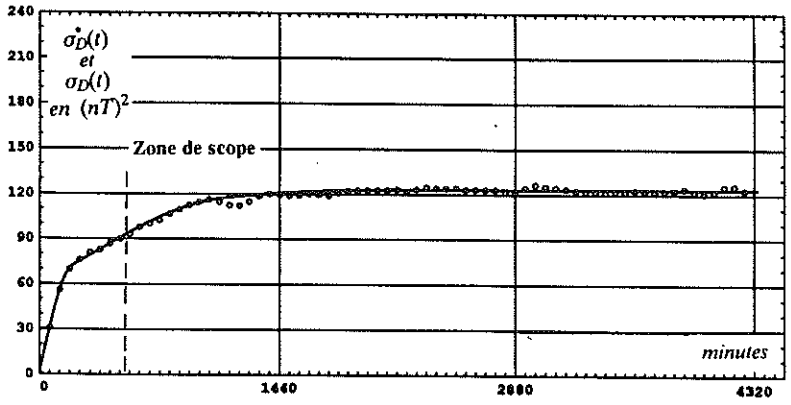


Figure 2.11 : Ajustement de la composante non périodique

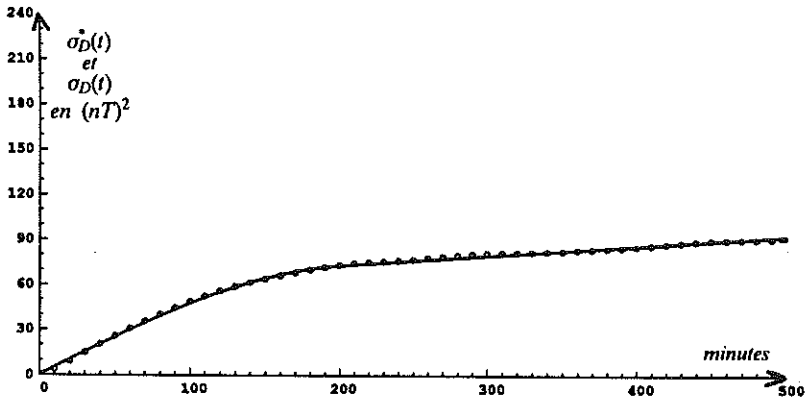


Figure 2.12 : Agrandissement des 500 premières minutes de l'ajustement

L'ajustement est effectué sur les 3 premiers jours de $\sigma^*_D(.)$, et les figures 2.11 et 2.12 permettront de juger sa qualité.

Etape 4, modéliser γ^*_D :

Les ajustements séparés de $\gamma_{\text{pério}}$ (par TFD) et de σ_D (manuellement) permettent ainsi de proposer un ajustement de γ_D , par la formule de reconstitution :

$$\gamma_D(h) = \sigma_D(h) + \gamma_{\text{pério}}(h) \ , \qquad \text{pour } h \in [0, 3 \text{ jours}] \tag{2.8}$$

avec:

$\sigma_D(.)$: somme des 3 structures à palier (étape 3).

$\gamma_{\text{pério}}(.)$: somme des 3 modèles cosinusoïdaux (étape 1).

Les figures 2.13, 2.14 et 2.15 permettent à trois échelles différentes de juger de la qualité de l'ajustement.

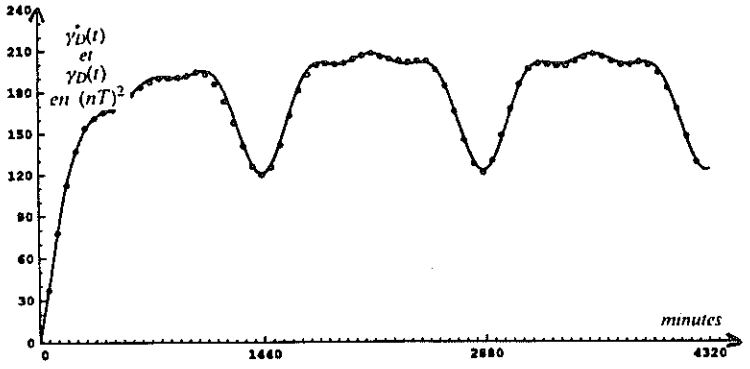


Figure 2.13 : 3 jours de l'ajustement du variogramme total (en trait continu : le variogramme expérimental, en trait pointé : le modèle)

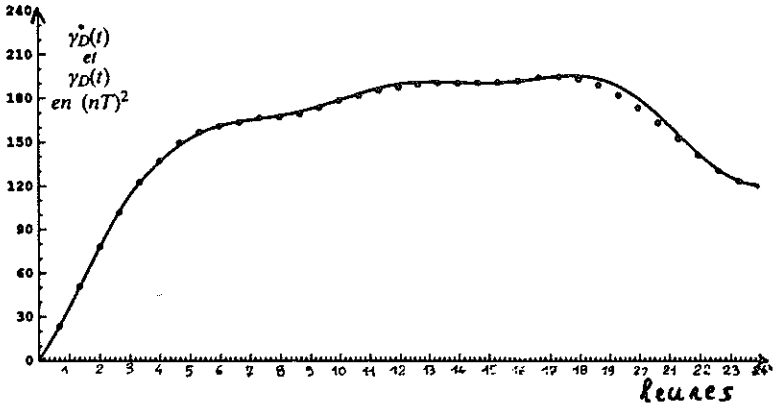


Figure 2.14 : Détail du premier jour

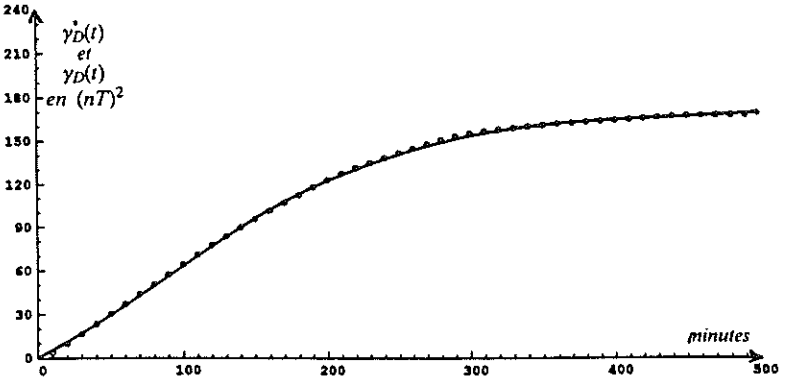


Figure 2.15 : Détail des 500 premières minutes

2 .2 .7 Conclusions sur Kakioka

Ce jeu de données représente une circonstance exceptionnelle par son champ (59 jours) et sa résolution extrêmement fine (une donnée par minute).
Cette circonstance a permis de cerner avec précision les caractéristiques de la composante périodique de D(.).

A cette circonstance s'ajoute celle du non-chevauchement des influences de la terre et de la lune, ce qui a permis de négliger cette dernière, au bénéfice d'un ajustement précis (mais sur 3 jours) du variogramme total γ_D . On ignore si cette absence de chevauchement constitue une loi physique propre aux perturbations diurnes, ou si elle est liée à l'implantation spatiale de la série temporelle et au domaine temporel analysé.

Cette section 2.2 se caractérise aussi par le fait qu'on ait constamment (et délibérément) occulté l'oscillation lunaire, en choisissant un champ de modélisation de γ^*_D égal à 3 jours. On peut s'interroger sur l'allure de la différence :

$$\gamma^*_{Lune} = \gamma^*_D - \gamma_D(h) \qquad (2.9)$$

pour h variant de ∅ à 25 jours avec, dans la formule 2.9, $\gamma_D(h)$ donné par la modélisation 2.8.

En théorie, ce variogramme résiduel devrait refléter l'influence périodique de la lune sur γ^*_D , et c'est pourquoi nous l'appelons γ^*_{Lune} .

Mais ce variogramme expérimental γ^*_{Lune} contiendra deux types d'imperfections :

– les imperfections liées aux simplifications inhérentes aux modèles $\gamma_{pério}$ et σ_D : en effet, lors de l'ajustement de $\gamma^*_{pério}$ par les seules périodes 24 h, 12 h et 8 h , on a laissé de côté près de 5 $(nT)^2$ d'énergie en n'utilisant que ces trois fréquences (tableau 2.4). Et de même, lors de l'ajustement de γ^*_D , on a occulté quelques micro-ondulations (figure 2.11)

Il en résulte que ce variogramme expérimental γ^*_{Lune} (qui, du fait de 2.9 est déjà un modèle) devrait être fortement bruité par les manipulations précédentes (bruité en ce sens qu'il doit contenir de l'énergie qui n'est pas nécessairement due à la lune).

– les autres imperfections liées à l'identification de γ_{Lune} par γ^*_{Lune} sont d'ordre statistique : en effet, comme la période de révolution de la lune autour de la terre est proche de 28 jours, et comme le champ d'analyse de D(.) est de 59 jours, cela représente un effectif statistique d'un peu plus de 2 (en termes de périodes) ce qui est nettement insuffisant (la pratique montre qu'un effectif d'au moins 7 périodes T est nécessaire pour bien appréhender variographiquement une onde de période T). On peut donc s'attendre à des dégénérescences .

La figure 2.16 qui contient les évolutions de γ^*_{Lune} (h) pour h variant de 0 à 26 jours, confirme nos présomptions.
En effet, on y observe une courbe approximativement sinusoïdale, de période 25 jours (et non 28) qui porte un signal périodique de période 24 h.

On retrouve là très exactement les conséquences des imperfections précédemment mentionnées, à savoir :

– une sous-évaluation de la période de la lune liée à un effectif de périodes trop faible

– un bruit non périodique lié aux modélisations de σ_D et $\gamma_{pério}$.

– un bruit périodique de période 24 h lié à une extrapolation excessive de la permanence temporelle de $\gamma_{pério}$.

Nous n'avons pas d'ajustement à proposer pour γ^*_{Lune} , bien que malgré ces handicaps, celui-ci eût été possible.

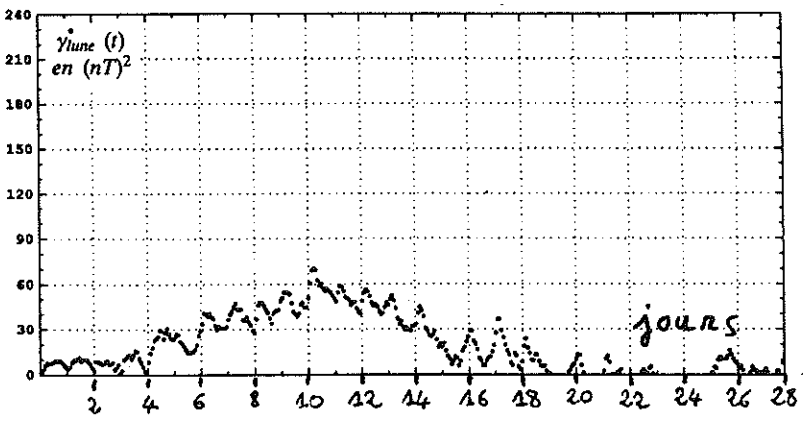


Figure 2.16 : 26 jours du variogramme de la composante lunaire

2 .3 Série temporelle à petite échelle d'analyse : Huancayo

2 .3 .1 Présentation des données

Ces données ont été enregistrées à l'observatoire fixe de Huancayo, au Pérou ($\simeq$ 20º N, 70º W). Elles nous ont été aimablement fournies par L. Ocola (réf [I-2]).
Elles sont à pas de 1 minute et mesurées du 22/8/1986 au 28/8/1986. Ce qui fait un total de 8640 temps informés. Ce champ temporel est donc approximativement 10 fois plus restreint que celui de Kakioka.

Comme précédemment, le signal a été translaté sur l'origine, et représente la mesure en nT de l'amplitude du vecteur champ magnétique total à l'observatoire.

2 .3 .2 Courbe expérimentale de D(t)

Les évolutions dans le temps de cet enregistrement ont déjà été présentées dans le chapitre 1.
La figure 1.3 montrait un processus possédant une composante périodique importante. On y constatait que les multiples de 12 ^{h} locales (soleil au zénith) constituaient les instants où il est le plus probable que D(t) atteigne ses maxima locaux.

Le domaine de variation de D(.) est de + 110 nT à - 100 nT, soit une amplitude crête à crête de 210 nT. Cette quantité est plus importante qu'à Kakioka parce que Huancayo est plus proche de l'équateur de 20 degrés de latitude.

Si on mesure, pour chaque période consécutive, le temps qui sépare les extrema, on trouve :

- Pour les maxima : une quasi-période qui varie de 1385 à 1456 minutes.
- Pour les minima : une quasi-période qui varie de 1410 à 1466 minutes.

Ceci donne une approximation de 1436 minutes pour la période moyenne .

2 .3 .3 Périodogrammes de D(t)

Comme cette zone d'étude ne contient que 6 périodes, nous avons effectué plusieurs périodogrammes, pour juger de la stabilité dans le temps, des amplitudes et fréquences émergentes. Nous présentons deux périodogrammes calculés respectivement sur les 3 premiers jours (figure 2.17, tableau 2.5) et sur les trois derniers jours (figure 2.18, tableau 2.6).

L'examen de ces résultats conduit à deux remarques essentielles :

- Comme pour Kakioka, les périodes 24 ^{h}, 12 ^{h}, et 8 ^{h} émergent des spectres d'amplitude, avec des coefficients relativement stables d'un périodogramme à l'autre.

  On a, en effet : pour le périodogramme 1 : $\simeq$ 38 nT, 20 nT, 7 nT pour respectivement 24 ^{h}, 12 ^{h} et 8 ^{h} ; et pour le périodogramme 2 : $\simeq$ 40 nT, 24 nT et 9 nT.
  Bien que ces fluctuations soient fortes dans l'absolu, elles peuvent être considérées comme modérées, compte tenu du fait que 3 périodes seulement sont utilisées pour chaque périodogramme.

- La constante de Fourier fluctue énormément puisqu'elle passe de $\simeq$ 0 nT (1er périodogramme) à - 15 nT (2e périodogramme).

  Ce résultat se comprend si on réexamine la figure 1.3.
  On y voit en effet qu'entre les jours 3 et 4, D(t) décroche brusquement vers le bas. Il en résulte un "niveau moyen", différent pour les 3 derniers jours.

Donc, bien qu'on dispose d'un champ temporel limité, on constate la permanence de l'émergence des périodes 24 ^{h}, 12 ^{h} et 8 ^{h}. Mais on ne peut nier d'importantes fluctuations d'un spectre d'amplitude à l'autre, en ce qui concerne la constante de Fourier.

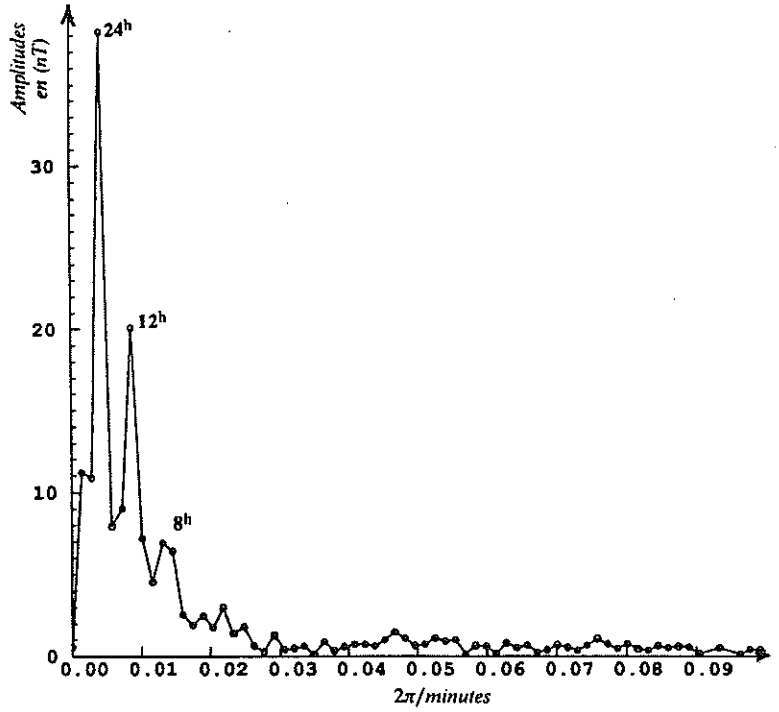


Figure 2.17 : Spectre d'amplitude des perturbations à Huancayo (calculé sur les 3 premiers jours de la série)

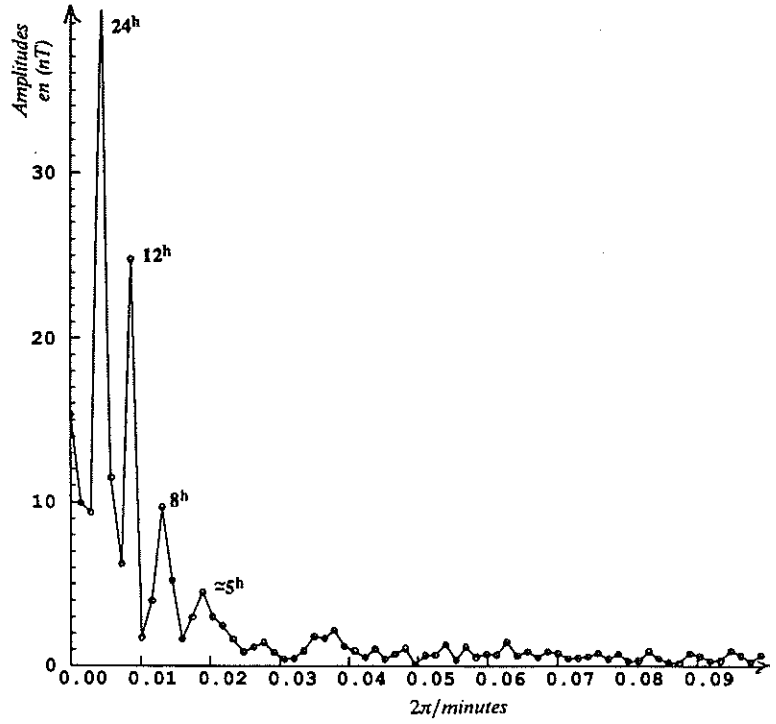


Figure 2.18 : Spectre d'amplitude des perturbations à Huancayo (calculé sur les 3 derniers jours de la série)

Fréquence ω_i	Périodes (minutes) T_i	Amplitude cosinus A_i	Amplitude sinus B_i	$\sqrt{A_i^2 + B_i^2}$	Phase (π radians) ϕ
0.0000000	0.00	-15.33	0.00	15.33	0.0000000
0.0014585	4308.00	6.79	7.27	9.95	0.2610417
0.0029170	2154.00	-8.63	3.66	9.38	0.8725305
0.0043755	1436.00	-18.46	-36.27	40.70	1.3501271
0.0058340	1077.00	8.47	-7.78	11.50	-0.2364168
0.0072925	861.60	1.76	6.02	6.27	0.4092917
0.0087510	718.00	-22.84	9.57	24.77	0.8737468
0.0102094	615.43	-1.74	0.09	1.75	0.9842818
0.0116679	538.50	-3.52	1.93	4.01	0.1594079
0.0131264	478.67	5.24	8.15	9.69	0.3179951
0.0145849	430.80	4.57	-2.54	5.23	-0.1616380
0.0160434	391.64	1.19	-1.16	1.66	-0.2468278
0.0175019	359.00	-2.43	-1.75	3.00	1.1991458
0.0189604	331.38	-3.98	-2.14	4.52	1.1573546
0.0204189	307.71	1.68	2.52	3.03	0.3128616
0.0218774	287.20	-0.44	2.44	2.48	0.5567935
0.0233359	269.25	0.67	-1.53	1.67	-0.3687887
0.0262529	239.33	-1.18	-0.05	1.18	1.0142726
0.0277114	226.74	0.27	1.45	1.47	0.4414831
0.0350038	179.50	1.81	-0.03	1.81	-0.0051483
0.0364623	172.32	-1.69	0.29	1.71	0.9454227
0.0379208	165.69	2.05	-0.84	2.21	-0.1246067
0.0393793	159.56	0.35	1.18	1.24	0.4086464
0.0437548	143.60	0.18	1.06	1.08	0.4451798
0.0481303	130.55	-0.51	0.98	1.10	0.6532655
0.0539642	116.43	-0.99	0.92	1.35	0.7632446
0.0568812	110.46	1.12	0.41	1.20	0.1125532
0.0627152	100.19	0.44	1.41	1.48	0.5966163

Tableau 2.5 : Amplitudes aux fréquences des perturbations (calculé sur les trois premiers jours)

Fréquence ω_i	Périodes (minutes) T_i	Amplitude cosinus A_i	Amplitude sinus B_i	$\sqrt{A_i^2 + B_i^2}$	Phase (π radians) ϕ
0.0000000	0.00	0.57	0.00	0.57	0.0000000
0.0014585	4308.00	3.45	-10.66	11.21	0.4005172
0.0029170	2154.00	3.06	-10.44	10.88	0.4091947
0.0043755	1436.00	-17.70	-33.90	38.24	1.3468475
0.0058340	1077.00	-1.20	-7.85	7.94	1.4516408
0.0072925	861.60	-8.14	-3.92	9.03	1.1427398
0.0087510	718.00	-15.51	12.79	20.11	0.7804852
0.0102094	615.43	-6.99	-1.80	7.22	1.0800372
0.0116679	538.50	-2.99	3.42	4.55	0.7288123
0.0131264	478.67	6.69	1.82	6.93	0.0845408
0.0145849	430.80	-0.41	6.43	6.44	0.5204872
0.0160434	391.64	0.43	2.53	2.57	0.4465203
0.0175019	359.00	-1.89	-0.46	1.94	1.0765240
0.0189604	331.38	1.43	2.08	2.52	0.3084100
0.0204189	307.71	0.90	1.54	1.79	0.3313782
0.0218774	287.20	0.68	2.94	3.02	0.4274958
0.0233359	269.25	-0.19	1.40	1.42	0.5419843
0.0247944	253.41	-1.23	1.38	1.84	0.7315531
0.0291699	215.40	0.57	1.19	1.32	0.3563688
0.0452133	138.97	0.90	-0.48	1.02	0.1556795
0.0466718	134.62	0.96	-1.15	1.50	-0.2782819
0.0481303	130.55	0.93	-0.63	1.12	-0.1909223
0.0525057	119.67	0.03	-1.13	1.13	-0.4907826
0.0554227	113.37	0.63	-0.79	1.01	-0.2872558
0.0758416	82.85	-0.45	1.01	1.10	0.6338403

Tableau 2.6 : Amplitudes aux fréquences des perturbations (calculé sur les trois derniers jours)

2 .3 .4 Variogramme expérimental de D(t)

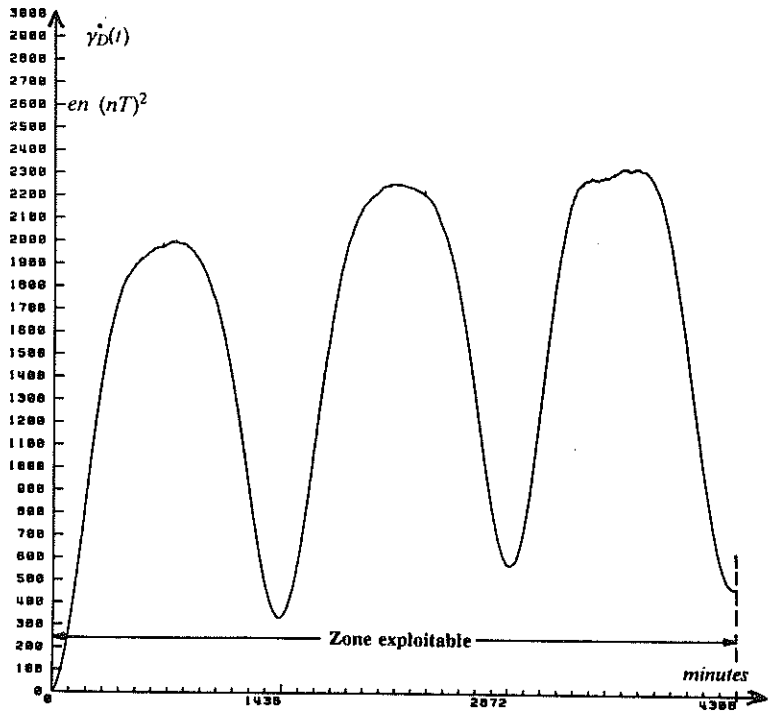


Figure 2.19 : Variogramme des perturbations

Bien que le champ soit restreint pour h = 3 jours, $\gamma_D^*(h)$ a été calculé à l'aide de près de 4 000 incréments, ce qui est encore statistiquement suffisant. Le variogramme est donc analysable jusqu'à cette distance. On présente sur la figure 2.19 les 3 premiers jours de $\gamma_D^*(.)$. On y note que $\gamma_D^*(.)$ présente trois ondulations de période $\simeq$ 24 h, d'amplitudes importantes, et de forme quasiment identique. On constate que ces ondulations ne reposent pas sur un palier, comme ce fut le cas pour Kakioka. Ceci nous obligera à n'ajuster $\gamma_D^*(.)$ que sur les 24 premières heures et rend inopérante la précédente méthode appliquée à Kakioka.. En conséquence, nous proposerons deux autres modes d'ajustement de cette fonction expérimentale.

Si on examine les périodes qui caractérisent ces ondulations, on constate qu'elles sont de 1420 minutes et non 1440 comme le laisserait supposer la connaissance physique de D(.). Un champ d'analyse de 7 périodes seulement conduit donc à une estimation peu précise de cette période principale.

Quoi qu'il en soit, c'est ce variogramme qui représente notre réalité expérimentale, et nous prendrons comme telle cette période de 1420 minutes qui constituera dès lors, la base de nos calculs.

2 .3 .5 Spectre d'énergie de D(t)

Nous avons effectué une TFD. de $\gamma_D^*(h)$, pour h variant de - 1420 à + 1420 minutes. On n'analyse donc que la première période de $\gamma_D^*(h)$.

Fréquence ω_i	Périodes (minutes) T_i	-Energie $\xi = TFD(\gamma_D^*)$
0.0000000	0.00	1345.64
0.0022124	2840.00	-56.45
0.0044248	1420.00	-856.15
0.0066372	946.67	-58.47
0.0088496	710.00	-259.40
0.0110619	568.00	-33.75
0.0132743	473.33	-45.84
0.0154867	405.71	-1.01
0.0176991	355.00	-5.17
0.0199115	315.56	-5.00
0.02212239	284.00	-2.21
0.0243363	258.18	-2.80
0.0265487	236.67	-1.44
0.0353982	177.50	-1.10
0.0376106	167.06	-1.60
0.0398230	157.78	-1.07
0.0442478	142.00	-1.04
0.0464602	135.24	-1.22
0.0530973	118.33	-1.15
0.0641593	97.93	-1.44

→ *Constante de Fourier*

Tableau 2.7 : Energies aux fréquences des perturbations (ne figurent que celles qui sont supérieures à 1 nT^2)

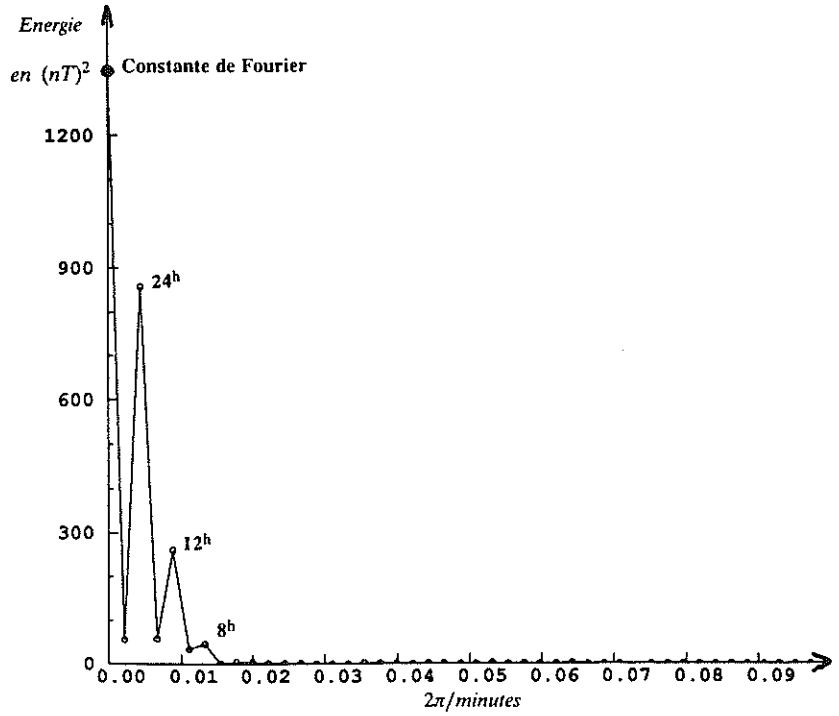


Figure 2.20 : Spectre d'énergie des perturbations

Hormis la constante de Fourier (positive comme il se doit), cette opération n'a engendré que des coefficients négatifs. Comme la période de base est ici de 1420 minutes, le tableau 2.7 ne fait apparaître que des diviseurs de cette quantité. Pour des raisons de clarté, nous continuerons à parler de 24 ^{h}, 8 ^{h} et 12 ^{h}, même si ce ne sont pas exactement ces valeurs qui entrent en jeu.. On note que celles-ci constituent la quasi-totalité du spectre. On comparera avantageusement les spectres Huancayo et Kakioka pour constater les importantes différences d'échelle de variance qui existent entre ces deux phénomènes, ainsi que leur rapport $\frac{\xi_{\text{pério}}(\emptyset)}{\sigma_D(\emptyset)}$

La constante de Fourier donne un estimateur de la variance totale:

$\xi_{\text{pério}}(\emptyset) + \sigma_D(\emptyset) \simeq 1345\ (\text{nT})^2$

2 .3 .6 Un premier ajustement : visuel

Si ce premier ajustement possède l'appellation de visuel c'est parce qu'il va être effectué en comparant "à vue" la courbe expérimentale γ^*_D aux modèles testés, suivant en cela la démarche géostatistique classique. Dans ce cas particulier de variogramme qui oscille, 2 ensembles de modèles sont considérés :

- les modèles à palier pour ajuster la composante non-périodique de γ^*_D (appelée σ_D),
- les modèles en cosinus pour la partie périodique.

Comme, par définition, σ_D est censé modéliser ce qui n'oscille pas dans γ^*_D , on s'attachera en premier lieu à ajuster ce qui oscille, c'est-à-dire à déterminer (sans TFD) les énergies et fréquences qui émergent dans γ^*_D . Contrairement aux pratiques habituelles, l'ajustement portera en premier lieu sur les grandes valeurs de h (ajustement de cosinus) puis ensuite seulement, sur le comportement de γ^*_D à l'origine (ajustement de σ_D).
Procédons par étapes successives.

Etape 1, période principale et palier de σ_D :

On choisit donc de limiter l'ajustement à la première période de γ^*_D . Ce choix équivaut à poser que la valeur de la période principale est de 1420 minutes (voir figure 2.21). Il en résulte une valeur du palier de σ_D (la structure non périodique) qui vaut :

palier $(\sigma_D) = \gamma^*_D$ (T = 1420 minutes) = 333 $(\text{nT})^2$

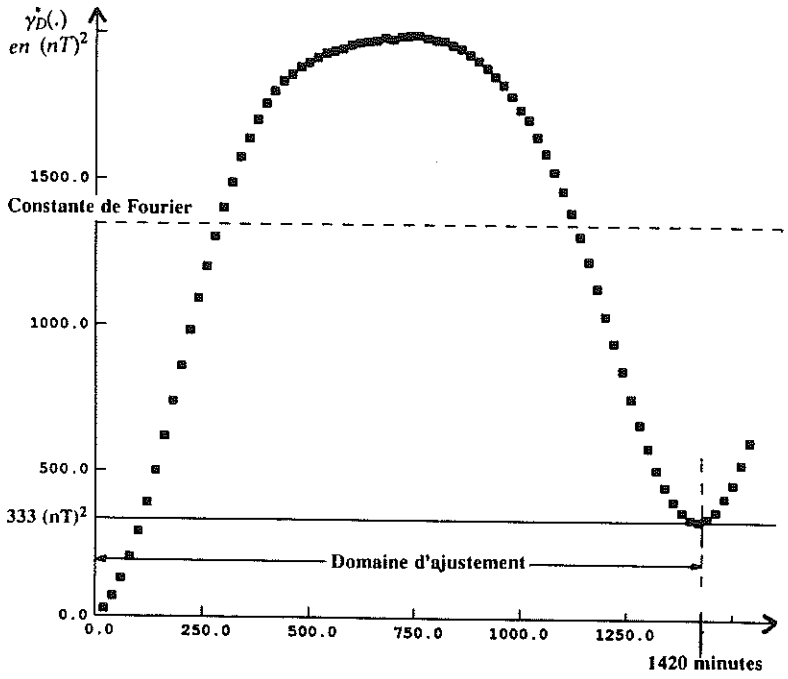


Figure 2.21 : Choix du domaine d'ajustement et détermination du palier

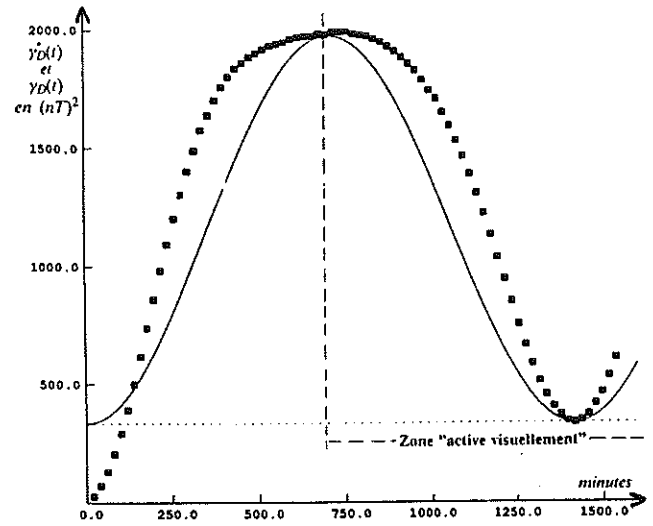


Figure 2.22 :
Ajustement de la période principale
(Trait pointillé : expérimental,
trait continu : le modèle)

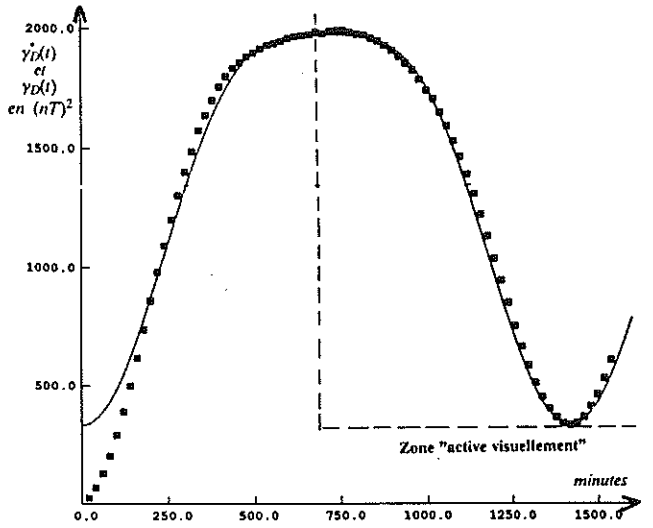


Figure 2.23 :
Ajustement de la période secondaire

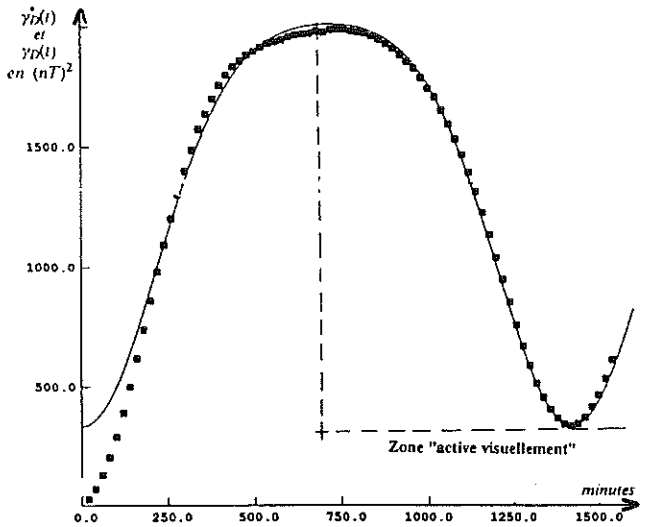


Figure 2.22 :
Ajustement de la période tierce

Etape 2, traitement de la composante périodique :

Notre champ d'analyse étant l'intervalle [0, 1420 minutes], un autre choix est fait: celui que σ_D possède une portée inférieure à $\frac{T}{2}$. Ce choix est discutable mais nécessaire si on souhaite ajuster manuellement des fréquences. En conséquence, on va momentanément considérer que σ_D est un effet de pépite de coefficient égal à 333 nT, pour s'appliquer à ajuster des modèles cosinus sur $\gamma_D^*(h)$, avec h variant de 710 à 1420 minutes (zone où σ_D n'est plus censé agir).

Comme ajuster un modèle périodique, c'est associer une énergie $\xi(\omega_i)$ à chacune des fréquences pressenties, nous procédons par étape :

- **La période principale (figure 2.22)**

  Elle résulte de l'étape précédente et vaut 1420 minutes.

  Il reste à déterminer son énergie. La figure 2.22 permet de comprendre que cette valeur ne peut être quelconque. La valeur obtenue par ajustement "visuel" est :

  $\xi(\omega_i) = 820(nT)^2$

- **La seconde période (figure 2.23)**

  La symétrie que $\gamma_D^*(h)$ possède de part et d'autre de h = 710 minutes oblige à prendre:

  $T_2 = 710$ minutes ,

  et la détermination de l'énergie associée à T_2 qui est :

  $\xi(\omega_2) = 200\ (nT)^2$,

  possède une fois encore quelque degré de généralité, si on souhaite préserver la forme curviligne du "dôme" de γ_D^*

- **Une troisième période (figure 2.24)**

  Ce n'est que par purisme que nous avons décidé d'ajuster une troisième période, car au stade de la figure 2.23, on pourrait estimer que le contenu périodique de γ_D^* est épuisé.

  La quantification de cette dernière période est plus délicate car il reste peu d'énergie à consommer. Nous nous sommes donc servi de l'expérience Kakioka pour choisir :

  $T_3 = 8$ heures $= 480$ minutes ,

  l'énergie qui lui est associée est :

  $\xi(\omega_3) = 20\ (nT)^2$

Le résultat de l'ajustement apparaît sur la figure 2.24. On peut estimer que la partie périodique de γ_D^* est correctement ajustée .

Etape 3, ajuster σ_D :

Les trois périodes 1420, 710 et 480 minutes et leurs énergies étant acquises, il reste à compléter l'ajustement de γ_D^* en remplaçant la covariance pépitique par un modèle à palier σ_D .

Plusieurs modèles étaient possibles, et nous avons finalement adopté un modèle prismato-magnétique (voir annexe B) de palier 333 $(nT)^2$ et de paramètre 150. L'ajustement qui en résulte est représenté sur la figure 2.25. Il peut être qualifié de satisfaisant, pour h variant de $\emptyset$ à 1420 minutes.

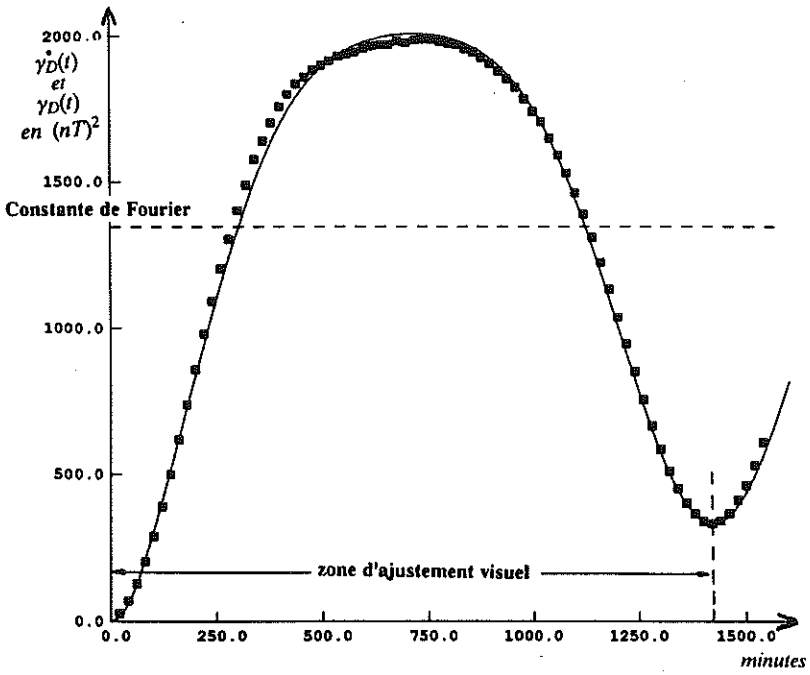


Figure 2.25 : Résultat de l'ajustement visuel du variogramme total (Trait continu : le modèle, trait pointillé : l'expérimental)

En conclusion, il se dégage de ce travail deux périodes (1420 et 710 minutes), comme pour Kakioka. Mais à l'inverse de Kakioka où la longueur du champ d'analyse était suffisante pour qualifier "d'expérimentales" les périodes obtenues, elles sont ici quelque peu arbitraires puisque déduites d'un champ d'analyse choisi. Il semblerait cependant que ces deux périodes soient caractéristiques des perturbations diurnes (d'autres jeux de données ne figurant pas dans cette thèse l'ont montré).

Il est intéressant de comparer les énergies obtenues manuellement (sur la composante périodique) à celles obtenues par la TFD de $\dot{\gamma}_D$ (tableau 2.7). On constate que :

- $\xi_{\text{manuel}}(1420) = 820\ (\text{nT})^2$ et $\xi_{\text{TFD}}(1420) \simeq 856\ (\text{nT})^2$
- $\xi_{\text{manuel}}(710) = 200\ (\text{nT})^2$ et $\xi_{\text{TFD}}(710) \simeq 259\ (\text{nT})^2$
- $\xi_{\text{manuel}}(480) = 20\ (\text{nT})^2$ et $\xi_{\text{TFD}}(473) \simeq 46\ (\text{nT})^2$

On constate que tous les $\xi_{\text{manuel}}(.)$ sont inférieurs aux $\xi_{\text{TFD}}(.)$ ce qui se comprend aisément puisque les ξ_{TFD} contiennent une partie de l'énergie de σ_D . Les deux approches sont donc compatibles, ce qui (d'un point de vue méthodologique) n'était pas évident a priori.

C'est en ce sens que nous dirons que les deux premières fréquences (et leurs énergies) sont des paramètres robustes : plusieurs méthodes différentes permettent d'aboutir au même résultat.

La manière dont nous venons de procéder n'est pas sans rappeler les techniques d'inférence d'une FAI-k décrites en annexe B.

En effet, il se dégage des 3 étapes précédemment étudiées le point suivant : nous avons dans un premier temps, ajusté 3 fréquences, en remplaçant la composante σ_D (inconnue) par un effet de pépite. Si une telle substitution s'est avérée opérationnelle, c'est parce que σ_D possédait une portée pratique égale à 350 minutes, soit une valeur inférieure à la plus petite des 3 périodes détectées ($\simeq$ 480 minutes) Cette circonstance a permis de modéliser la tendance à grande échelle du processus..

Lorsqu'on souhaite appréhender une FAI-k , et déterminer son degré k, on remplace momentanément cette dernière par un effet de pépite et ceci jusqu'à la détermination du degré "k".

On voit donc qu'on se trouve dans une circonstance similaire où remplacer momentanément σ_D par l'effet de pépite a permis d'évaluer (en termes de fréquences et énergies) une "tendance" qui s'apparente à un degré "k", mais qui porte sur des caractéristiques spectrales.

2 .3 .7 Un second ajustement : par TFD

Ce deuxième ajustement est immédiat. Il consiste en l'application des opérations suivantes :

- Le calcul d'une TFD du variogramme pour h variant de -24 ^{h} à 24 ^{h} (tableau 2.7).
- Une sélection des fréquences à forte énergie (tableau 2.8).
- Une compensation de l'énergie laissée de coté (ici 33 nT^2) par un effet de pépite.

Fréquence ω_i	Périodes (minutes) T_i	-Energie $\xi = TFD(\gamma^*_D)$
0.0022124	2840.00	-56.45
0.0044248	1420.00	-856.15
0.0066372	946.67	-58.47
0.0088496	710.00	-259.40
0.0110619	568.00	-33.75
0.0132743	473.33	-45.84

Tableau 2.8 : Fréquences et énergies sélectionnées pour l'ajustement

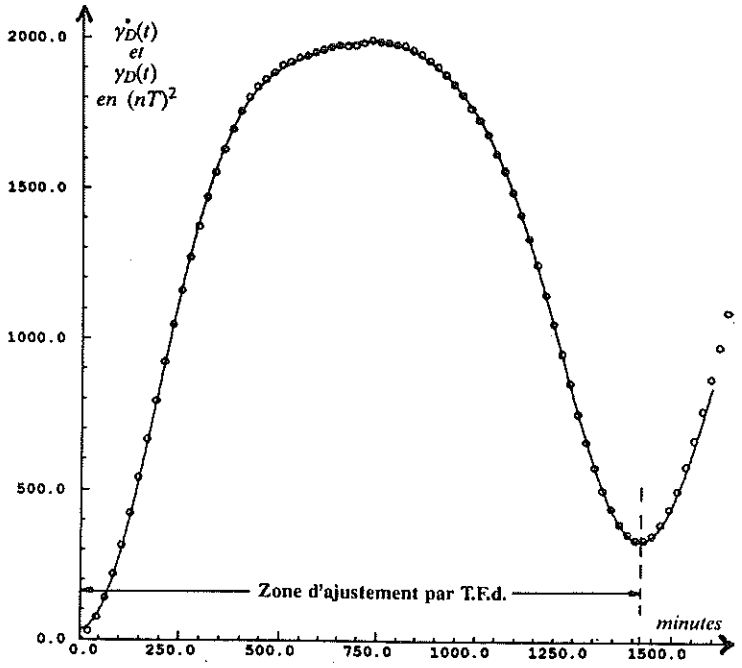


Figure 2.26 : Résultat de l'ajustement du variogramme total par T.F.D. (Trait continu : le modèle, trait pointillé : l'expérimental)

La figure 2.26 permet de juger de la qualité de l'ajustement résultant.

On notera que la coupure spectrale utilisée portant essentiellement sur les hautes fréquences, cet ajustement ne respecte pas les 15 premières minutes de γ_D^* .

En fait, on aurait pu utiliser un modèle à palier de portée extrêmement faible, en remplacement de l'effet de pépite, ce qui aurait permis d'ajuster parfaitement l'origine de γ_D^*. C'est par pur souci d'économie de modèle que nous ne l'avons pas fait ici.

2 ,3 ,8 Conclusions sur Huancayo

Bien que ce jeu de données soit (en terme de périodes) peu étendu, il aura cependant permis une modélisation correcte, mais dont l'utilisation doit être limitée à un champ plus restreint que pour Kakioka. Le fait que ce champ restreint ait dégagé une période principale de 1420 minutes seulement (et non 1440) montre qu'un champ couvrant plus de périodes doit être analysé pour appréhender convenablement ce PTP.

L'inexistence, dans γ_D^*, d'une zone de transition telle que celle rencontrée dans Kakioka a conduit à deux méthodes d'ajustement :

- L'une procédant d'une dichotomie en termes périodique et non-périodique et en un ajustement manuel d'énergies.
- L'autre procédant d'une TFD et d'un filtre fréquentiel.

Ces deux méthodologies sont très différentes quant à leur modalité de mise en oeuvre, mais similaires par la qualité du résultat qu'elles engendrent.

2 .4 Série temporelle à faibles oscillations : MOHNS

Cette troisième série temporelle se caractérise par son implantation spatiale très proche du pôle Nord. Ceci implique (cf section 2.1.1) que les oscillations sont peu marquées et se prêtent difficilement à une modélisation périodique. Ce dernier exemple va donc permettre de montrer les limitations de notre approche lorsque les perturbations diurnes présentent de telles caractéristiques.

2 .4 .1 Présentation des données

Contrairement aux deux cas précédents, cette série n'est pas issue d'un observatoire terrestre, mais obtenue à l'aide d'un magnétomètre immergé. Ces mesures ont été faites à partir du 9 Juillet 1988, lors de la campagne océanographique MOHNS (4-27 Juillet 1988), par 2 400 mètres de fond.
Ces données nous ont aimablement été fournies par V. Renard (ref [I-3]).

Lors de cette campagne MOHNS-88 à laquelle P. Chauvet (ref [I-1]) et moi-même avons participé, l'un des objectifs était d'engendrer un jeu de données spatio-temporel permettant de tester certaines méthodes géostatistiques, et notamment l'estimation des perturbations diurnes à partir du magnétisme spatio-temporel.
Dans ce but, nous avions immergé un enregistreur dont la courbe servirait de vérification à nos calculs.
Pour des raisons inconnues, celui-ci, bien que perfectionné, n'a fonctionné que 3 jours pour une campagne qui en durait 15. Le froid (immersion à -100 mètres), la salinité des eaux, l'immersion elle-même, sont des contraintes extrêmement rudes pour un appareillage aussi sensible. Il en résulta que nous ne pûmes, lors de cette campagne, corriger les données à l'aide de ces enregistrements et que nous ne disposons que de 3 jours de mesure.

La localisation du capteur immergé était : Latitude N 72º, longitude W 2º , soit approximativement à 700 km à l'Est de l'Ile Jean Mayen.

2 .4 .2 Courbe expérimentale de D(t)

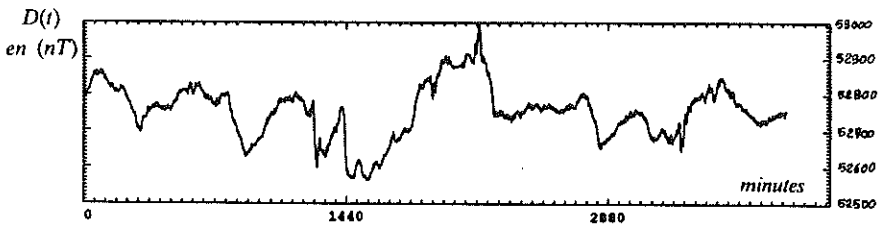


Figure 2.27 : Evolutions dans le temps du champ magnétique total au capteur immergé à Mohns (cercle polaire)

La courbe 2.27 montre les évolutions dans le temps des 3 jours de mesures prélevées par pas de 2 minutes. Comparativement aux séries KAKIOKA et MOHNS, deux remarques s'imposent :

- l'amplitude maximale crête à crête du signal est de $\simeq$ 400 nT, alors qu'elle était de 100 nT pour Kakioka et $\simeq$ 210 nT pour Huancayo.
- visuellement, on ne perçoit pas de périodicité.

Ces remarques jointes à la localisation boréale de l'appareil de mesure, corroborent ce qui a déjà été dit sur les perturbations diurnes en section 2.1, à savoir que ce phénomène perd de son intensité relative, et sa structure périodique, lorsqu'on approche les pôles.

2 .4 .3 Périodogramme de D(t)

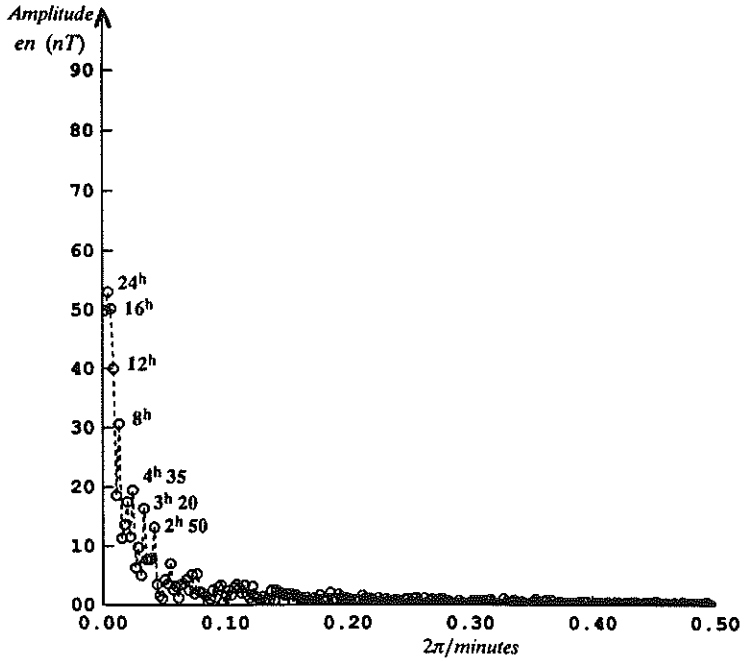


Figure 2.28 : Spectre d'amplitude des perturbations

Fréquence ω_i	Périodes (minutes) T_i	Amplitude cosinus A_i	Amplitude sinus B_i	$\sqrt{A_i^2 + B_i^2}$	Phase (π radians) ϕ
0.0000000	0.000	52757.4	0.000	527.57.4	0.000
0.0021817	2880.000	45.2	-21.3	49.9	-0.140
0.0043633	1440.000	-51.9	10.8	53.0	0.935
0.0065450	960.000	30.4	40.0	50.2	0.293
0.0087266	720.000	7.9	-39.3	40.0	-0.437
0.0109083	576.000	6.4	17.5	18.6	0.389
0.0130900	480.000	-17.2	25.4	30.7	0.689
0.0152716	411.429	-9.4	06.1	11.3	0.817
0.0174533	360.000	11.2	07.7	13.6	0.193
0.0196350	320.000	16.9	04.5	17.5	0.916
0.0218166	288.000	-10.1	-05.6	11.5	1.162
0.0239983	261.818	-9.6	17.0	19.5	0.663
0.0327249	192.000	-1.2	16.4	16.4	0.524
0.0414516	150.579	-0.9	13.2	13.2	0.521

Tableau 2.9 : Amplitudes aux fréquences des perturbations

L'examen des figures et tableaux présentés ne permet pas de constater l'émergence des fréquences 24 ^{h}, 12 ^{h} et 8 ^{h}. Bien que la période 24 ^{h} possède une amplitude

relativement importante, celle-ci s'inscrit dans la continuité d'un spectre qui décroît rapidement vers ∅ (ie : modèle à covariance à palier) et ne procède pas d'une émergence. On ne reconnaît pas là le spectre d'un PTP examiné dans son espace-période.

On notera que les amplitudes sont faibles, relativement à l'amplitude crête à crête de 400 nT du signal qui laissait présager des amplitudes au moins égales à 100 nT. Ceci est dû au fait que l'énergie du signal se décompose sur pratiquement toutes les fréquences de la TFD. Ceci n'apparaît pas dans le tableau 2.9 (puisque celui-ci ne contient que les amplitudes supérieures à 10 nT), mais est clairement illustré par la figure 2.27. Autrement dit, si le spectre décroît rapidement vers zéro, il est cependant difficile de le considérer comme borné. Ce fait existait déjà dans les séries Huancayo et Kakioka, mais possédait une importance relative (à l'énergie totale) négligeable.

2 .4 .4 Variogramme de D(t)

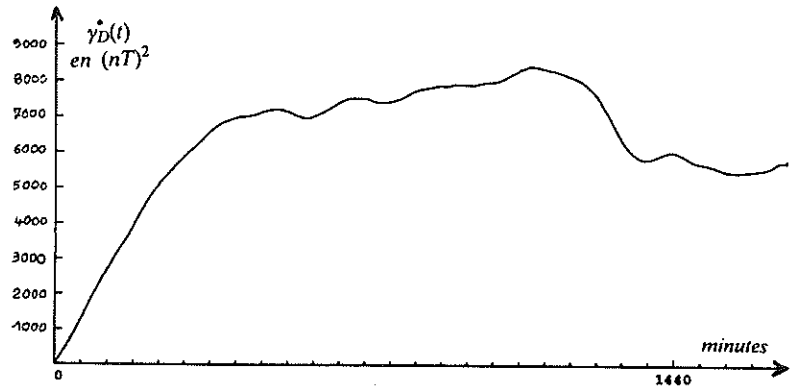


Figure 2.29 : Variogramme expérimental des perturbations

Ce variogramme a été calculé par pas de 2 minutes, avec une tolérance nulle, et n'est examiné que sur les 28 premières heures, en raison du champ temporel des données qui est limité à 56 heures. Son examen montre un comportement à l'origine et jusqu'à h $\simeq$ 8^h, identifiable à un modèle sphérique dont le palier serait de 7 000 $(nT)^2$ et égal à la variance expérimentale (centrée) des données. Si visuellement, on perçoit quelques oscillations, celles-ci sont peu structurées et rendent inapplicables les méthodes précédentes d'ajustement. Son éventuel ajustement relève donc de la Géostatistique classique, et pour cette raison, n'est pas présent ici.

Moyennant un certain empirisme , et en s'aidant de la TFD du tableau 2.9, on peut imaginer d'ajouter au σ_D sphérique précédemment pressenti un certain nombre de cosinus, afin de respecter avec exactitude $\gamma^*(.)$.

2 .4 .5 Spectre d'énergie de D(t) et conclusions

L'examen du spectre d'énergie (figure 2.30) n'apporte rien de nouveau à celui des amplitudes. Une manière de conclure brièvement sur la série MOHNS, serait de dire que si son étude fut si courte, c'est parce qu'il ne s'agit pas d'un PTP. !En effet, force est de constater que le spectre de la figure 2.8, caractéristique de nos PTP, ne possède aucun lien de parenté avec celui rencontré ici..

L'intérêt de ce cas d'étude est d'ordre naturaliste et méthodologique, car il montre les limites régionales dans lesquelles notre manière d'appréhender les oscillations du magnétisme est opérationnelle. Ces limites peuvent être évaluées à l'intervalle [-65 º + 65 º] en latitude. Dans cette bande spatiale, les perturbations ont une tendance périodique et peuvent être appréhendées comme nous venons de la faire. En dehors de cet intervalle, c'est la géostatistique classique (et éventuellement stationnaire) qui répondrait au problème de leur analyse et modélisation.

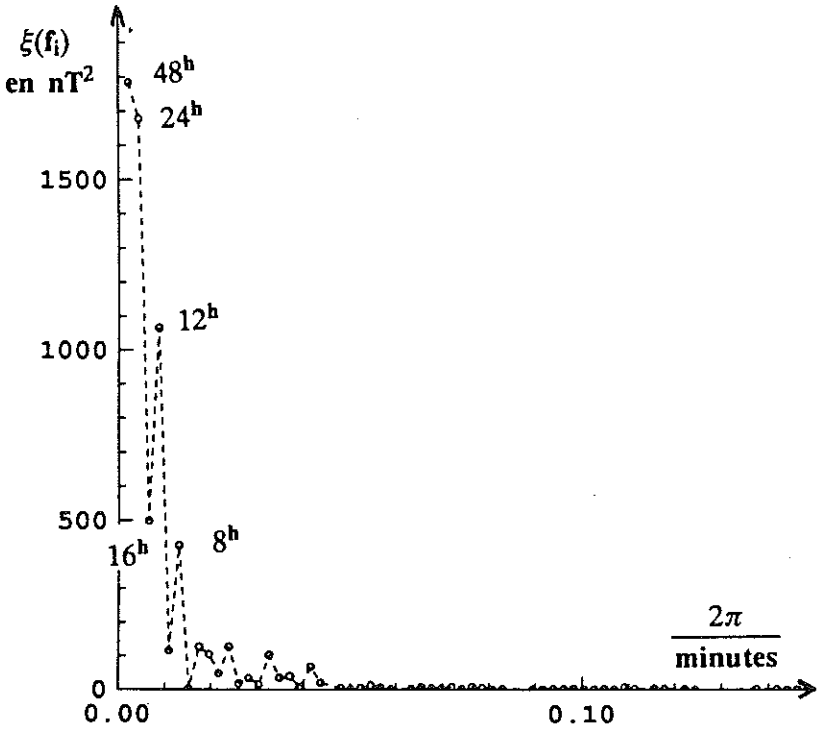


Figure 2.30 : Spectre d'énergie des perturbations

2 .5 Conclusions du chapitre 2

La manière dont l'analyse de ces séries vient d'être conduite montre la complémentarité des outils spectraux et spatiaux.

Lorsque les conditions expérimentales ne permettent pas d'extraire du variogramme expérimental sa composante périodique, et de ce fait de l'analyser par TFD (exemple Huancayo), l'ajustement manuel que nous avons pratiqué alors s'est avéré également très efficace, tant par la précision de l'ajustement résultant que par la quantification des énergies des oscillations du PTP. Ceci a permis de qualifier de "robuste" les paramètres "période et énergie " du PTP.

Il se dessine une pratique généralisable à des VR autres que le PTP: celle d'examiner simultanément le spectre d'énergie et le variogramme de la VR analysée.

L'un des premiers avantages d'une telle 'vision binoculaire' du phénomène est de permettre de choisir entre deux modèles qui dans l'espace donnent apparemment une bonne interpolation du variogramme expérimental. C'est le cas de l'exemple Huancayo où il subsiste une alternative: choisir entre le modèle obtenu par TFD pure et celui obtenu manuellement. Dans les 2 cas, ces modèles respectent l'expérimental avec peu de moyens (6 fréquences et un effet de pépite pour l'ajustement par TFD, 3 fréquences et 1 structure pour l'ajustement manuel), de sorte que le principe d'économie de modèle ne peut être un critère de choix.

On comprendra donc que devant de telles alternatives, il peut s'avérer utile d'examiner les spectres d'énergie associés aux modèles et de les comparer au spectre expérimental. Respecter également l'allure du spectre (et pas seulement celle du variogramme) permettrait de choisir entre ces modèles équivalents.

Ainsi, il est aisé d'imaginer (mais sans doute lourd de développer) un programme d'ajustement, tel qu'il est aujourd'hui, mais auquel s'ajouterait une fenêtre de visualisation spectrale,

Chapitre 3

ANALYSE DE PTP SPATIO-TEMPORELS (magnétisme)

Dans ce chapitre, on s'attache à analyser des champs de séries temporelles de magnétisme terrestre, chacune d'elle étant une série de perturbations diurnes. Ce PTP est caractérisé par son espace de régionalisation qui est le plan-temps, et par son espace-période qui est le temps. Le fait que cette VR soit mesurée par un bateau se déplaçant sur la zone d'étude nécessitera certaines hypothèses sur la variable, des configurations particulières de voisinage d'analyse, et la modification des boucles d'inférence de la FAI-k.
L'ensemble de ces spécificités fait l'objet de la section 3.1.

Ces méthodes seront ensuite mises en oeuvre sur les données de la campagne océanographique KRAKATAO (section 3.2), SEAPERC (section 3.3), et MOHNS (section 3.4). Comme pour SEAPERC et MOHNS, nous disposons d'enregistrements de perturbations diurnes faits sur zone et étudiés dans le chapitre 2, il sera possible de juger de la qualité des analyses.

Ce chapitre permettra également d'introduire les modèles qui serviront dans le chapitre 4 pour des opérations d'estimations.

3 .1 Spécificités du problème

Dans cette section, nous présentons la manière dont seront conduites l'analyse et la modélisation du PTP "magnétisme spatio-temporel". Après avoir présenté la gène causée par les perturbations diurnes lors de l'analyse de données géomagnétiques, nous insisterons sur les problèmes posés lorsque les données sont alignées sur des profils sensiblement parallèles.
Nous montrerons ensuite comment modifier les boucles d'inférence des paramètres pour mener à bien l'analyse et la modélisation de ce PTP.
Enfin, nous présenterons les 4 géométries qui seront utilisées pour le voisinage d'analyse (noté $\mathcal{V}$).

3 .1 .1 Le problème des perturbations diurnes en Géophysique marine

L'étude des anomalies du magnétisme de l'écorce terrestre est essentielle en Géophysique marine, car elle permet de connaître l'histoire de l'ouverture des dorsales médio-océaniques et révèle la nature des fonds marins.
Or le champ géomagnétique dépend aussi de causes externes à l'origine de perturbations diurnes, mensuelles et séculaires. Il en résulte que lorsqu'un bateau effectue des mesures de magnétisme en se déplaçant sur la zone d'étude, ces données ne sont pas directement analysables, puisque brouillées par ces perturbations.

En Géophysique Marine, seules les perturbations diurnes sont à supprimer des anomalies magnétiques, du fait que les campagnes de mesure sont en moyenne d'une dizaine de jours sur une même zone. Les perturbations lunaires ne sont pas alors décelables.
Ces perturbations impliquent qu'aucune analyse de données magnétiques dans les régions équatoriales ne peut se passer de prendre en compte les perturbations diurnes (réf [GΦ-2]). De plus, comme les anomalies magnétiques océaniques sont faibles près de l'équateur, le rapport signal/bruit est petit, ce qui rend impossible le tracé d'isovaleurs et difficile l'interprétation des anomalies.

Que faire alors ? Une première possibilité consiste à disposer d'un enregistreur fixe à terre (près de la zone d'étude) pour enregistrer le champ géomagnétique et son évolution dans le temps, permettant ainsi de corriger les mesures spatio-temporelles de la campagne.

C'est la pratique la plus courante, et qui est à l'origine de la série temporelle Huancayo (correction de la campagne Seaperc).

Mais cette solution s'avère abusive lorsque l'observatoire se situe trop loin de la zone d'étude, car dans ce cas, les fluctuations temporelles peuvent différer sensiblement de celles de la zone d'étude.

Une autre possibilité consiste à immerger (au début de la campagne) un magnétomètre enregistreur en bordure de la zone d'étude. C'est le cas de la série Mohns (correction de la campagne Mohns). Mais peu de chefs de mission le font, pour des raisons de gain de temps, de fiabilité d'appareillage, et du fait que bien souvent lorsque la campagne s'achève et qu'il s'agit de récupérer l'appareillage immergé, celui-ci a mystérieusement disparu ...

Une dernière possibilité consiste à estimer ces perturbations à partir des données spatio-temporelles. C'est l'un des objectifs de cette thèse. Comme une telle estimation suppose une modélisation de la VR, il convient au préalable d'analyser de tels phénomènes.

3 .1 .2 Particularités de cet échantillonnage spatio-temporel

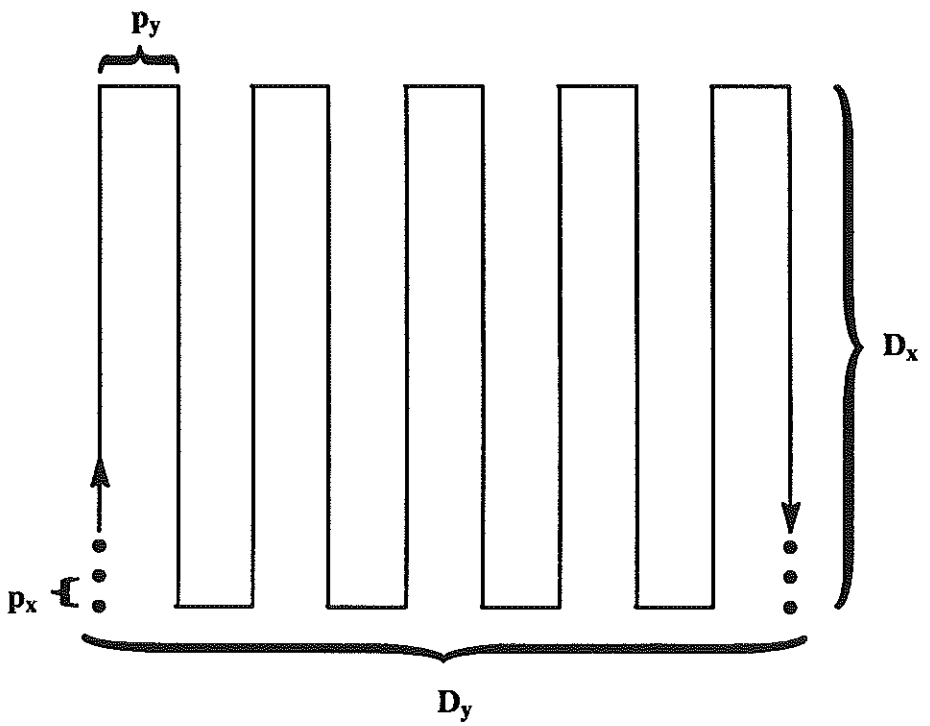


Figure 3.1 : Route d'échantillonnage en Géophysique Marine

Route du bateau :

La route du bateau est bien souvent constituée de profils parallèles équidistants. Les données sont à pas de 1 minute de navigation le long des profils et les profils peuvent être distants de plusieurs km, de sorte que l'échantillonnage est anisotrope dans un rapport pouvant aller de 10 à 100. Il est alors possible de définir des vecteurs pas et diamètres d'échantillonnage) :

(D_x, D_y) : vecteur diamètre de la zone d'étude (3.1)

(p_x, p_y) : vecteur pas d'échantillonnage du bateau (3.2)

La figure 3.1 montre cette configuration typique.

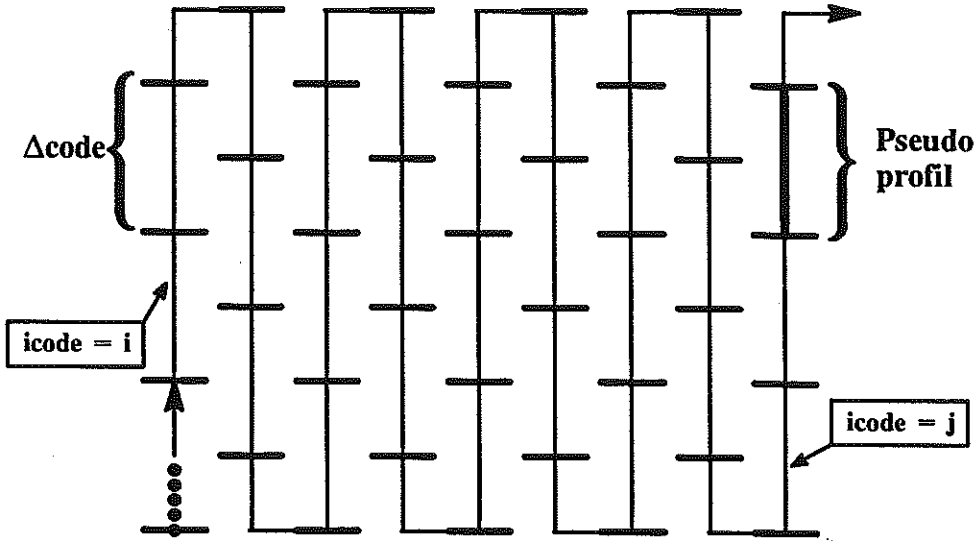


Figure 3.2 : Décomposition de la route en "pseudo-profils"

Notion de "pseudo-profil" :

Bien que nous disposions de données sur profil, cette anisotropie de l'échantillonnage amène à effectuer une décomposition de la route en "pseudo-profils", ceci afin de pouvoir mieux gérer l'information. Un pseudo-profil sera pour nous un paquet d'informations successives, identifié par une étiquette entière appelée icode et caractérisé par sa longueur Δcode :

icode : étiquette du pseudo-profil. (3.3)

Δcode : longueur du pseudo-profil. (3.4)

Comme nous serons amené à travailler en voisinage glissant (voir annexe B-4), nous aurons aussi à définir le nombre de pseudo-profils contenus dans chaque voisinage. Nous désignerons par Ncode ce nombre:

Ncode : nombre de pseudo-profils par voisinage d'analyse. (3.5)

La figure 3.2 montre un exemple de telle décomposition de la route. Dans certains cas, la décomposition sera celle des véritables profils.

Echantillonnage quasiment coplanaire :

Si en toute généralité, le bateau échantillonne un champ de séries temporelles de perturbations diurnes dont chaque niveau de référence est l'anomalie spatiale, la manière dont sont conduits les prélèvements fait que nous ne disposerons en général que d'une valeur de la série pour une implantation spatiale donnée . Si donc on examine cette route dans l'espace temps, on constate que celle-ci est quasiment contenue dans un plan. Il en résulte que le "cube" spatio-temporel dans lequel le PTP est régionalisé est excessivement mal échantillonné. Ce fait rejaillira nécessairement dans la suite, à tous les niveaux de l'analyse et de la modélisation. De ce point de vue, un véritable champ de séries temporelles remplissant ce cube spatio-temporel représenterait une circonstance beaucoup plus favorable pour l'analyse.

3 .1 .3 Hypothèses

Afin de pouvoir appréhender ce PTP, certaines hypothèses sont nécessaires. Leur validité sera confirmée ultérieurement par la qualité des modèles inférés. Examinons ces hypothèses.

Dichotomie :

Appelons z(x,t) la VR magnétisme spatio-temporel, avec pour "x" le couple des coordonnées spatiales et pour "t" la localisation temporelle de la mesure. On interprètera z(x,t) comme la réalisation d'une FA Z(x,t) qui procède de la dichotomie suivante:

$$Z(x,t) = Y(x) + D(t) \tag{3.6}$$

avec :

$Y(x)$: anomalie spatiale (3.7)

$D(t)$: perturbations diurnes (3.8)

Ce modèle suppose que les perturbations diurnes sont indépendantes de leur implantation spatiale, ce qui n'est vrai que sur un domaine spatial limité. Cette hypothèse a été jugée acceptable sur les domaines couverts par les campagnes océanographiques, par les Géophysiciens consultés.

Modélisation de Y(x) par une FAI-k :

Les anomalies spatiales Y(x) peuvent être quelconques, et en particulier non nécessairement stationnaires. C'est pourquoi on choisit de les modéliser par l'outil le plus général que l'on possède: celui des Fonctions Aléatoires Intrinsèques d'ordre k (FAI-k, cf annexe B-3).

On supposera dans la suite que Y(x) est une FAI spatiale de degré inconnu "k" et de Covariance Généralisée (CG) inconnue K(h) ("h" désignera dans la suite une distance spatiale).

Modélisation de D(t) par une série trigonométrique :

D(t) étant noyée dans l'espace-temps, nous renonçons (dans le modèle) à sa composante non périodique (de covariance $\sigma_D(.)$, cf. chap. 2) et nous choisissons de l'identifier à une série de Fourier (S$\mathcal{F}$) :

$$D(t) = \sum_{i=\emptyset}^{Nfrèq} [\ A_i \cos(2\pi f_i t) + B_i \sin(2\pi f_i t)\ ] \qquad \forall\ t \tag{3.9}$$

Dans 3.9, les coefficients A_i et B_i (inconnus) seront supposés constants à l'échelle du voisinage d'analyse. Dans cette même formule, les fréquences f_i sont inconnues, mais la connaissance que nous avons de la physique de D(t) nous guidera quant aux choix possibles de celles-ci.

3 .1 .4 Modifications des boucles d'inférence

Notre travail va consister à déterminer, à partir d'une réalisation z(x,t) et pour une géométrie fixée du voisinage (que nous noterons $\mathcal{V}$), l'ensemble [k , $\{f_i\}$, K(h)] des paramètres caractéristiques de ce PTP. La présence du phénomène périodique D(t) nécessite de modifier les boucles d'inférences présentées en annexe B, pour tenir compte des spécificités du problème.Détaillons ces points.

Espace des CLA-k-$\{f_i\}$:

Comme les contraintes expérimentales empêchent de pouvoir examiner D(t) dans son espace-période, et de ce fait empêchent de pouvoir examiner Y(x) dans le plan seulement, vouloir atteindre les caractéristiques k et K(h) de Y(x) nécessite de travailler sur une variable spatiale uniquement. On choisit donc de travailler sur des combinaisons linéaires $\sum_\alpha \lambda^\alpha z(x_\alpha, t_\alpha)$, avec $\{\lambda_\alpha\}$ tels que:

$$\sum_\alpha \lambda^\alpha \cos(2\pi f_i t_\alpha) = \emptyset \qquad i : 1 \ldots Nfrèq \tag{3.10}$$

et :

$$\sum_\alpha \lambda^\alpha \sin(2\pi f_i t_\alpha) = \emptyset \qquad i : 1 \ldots Nfrèq \tag{3.11}$$

Nous appellerons CLA-$\{f_i\}$ des combinaisons linéaires telles que $\{\lambda_\alpha\}$ vérifie 3.10 et 3.11. Dans notre modèle, ces CLA-$\{f_i\}$ rendent l'expression $\sum_\alpha \lambda^\alpha z(x_\alpha, t_\alpha)$ indépendante du temps. Pour

atteindre K(h) , il reste à les rendre indépendantes des polynômes de degré inférieur ou égal à k. C'est pourquoi on imposera de plus aux coefficients $\{\lambda_\alpha\}$ d'être tels que :

$$\sum_\alpha \lambda^\alpha f_l(x_\alpha) = \emptyset \qquad \forall\ l \leq k \tag{3.12}$$

ie que $\sum \lambda^\alpha z(x_\alpha, t_\alpha)$ soit également une CLA-k (cf. annexe B). Dans cette expression, on rappelle que les fonctions $f_l(x)$ désignent des polynômes spatiaux en x et y de degré inférieur à k. Nous dirons alors qu'une combinaison linéaire $\sum \lambda^\alpha z(x_\alpha, t_\alpha)$ est une CLA-k-$\{f_i\}$ si $\{\lambda_\alpha\}$ vérifie 3.10 à 3.12, et nous noterons $\Lambda_{k-\{f_i\}}$ l'ensemble de ces coefficients.

En résumé: les particularités de ce PTP conduisent à orthogonaliser les mesures $\{\lambda_\alpha\}$ sur l'espace $\mathcal{F}$ suivant :

$$\mathcal{F} = \mathcal{F}_k \oplus \mathcal{F}_{\{f_i\}} \qquad \text{: somme directe} \tag{3.13}$$

avec :

$\mathcal{F}_k$: espace engendré par les polynômes spatiaux en x et y de degré $l \leq k$. (3.14)

$\mathcal{F}_{\{f_i\}}$: espace engendré par les cosinus et sinus temporels des fréquences . (3.15)

Principe de l'inférence :

A l'espace $\mathcal{F}$ près, l'inférence se pratiquera d'une manière comparable à celle d'une FAI-k classique. Simplement, les "boucles-k" (cf annexe B) deviendront des "boucles-k-$\{f_i\}$ ". Nous procèderons donc d'une manière classique, à savoir par une succession "d'essais-erreurs" sur les différents paramètres k, $\{f_i\}$, et K(h) . A ces boucles s'ajouteront celles d'estimations "in-situ", où le modèle sera directement testé sur les données. On parlera alors toujours de "boucle- K^* " (validation du modèle). Dans chacune de ces boucles, nous construirons des CLA-k-$\{f_i\}$ sur lesquelles nous confronterons hypothèse et réalité. L'inférence se fera toujours sur des "zones test" réparties d'une manière homogène sur le champ d'analyse .

Inférence de k et $\{f_i\}$ (boucle-k-$\{f_i\}$) :

Fixons nous sur l'une de ces zones test $\mathcal{V}(x_i, t_i)$ (étiquetée "ivois"), et faisons l'hypothèse que $k = k_{prior}$ et $\{f_i\} = \{f_i\}_{prior}$
(l'étiquette "prior" indique que la valeur à laquelle elle est attachée est arbitraire, momentanément on l'espère).

Après avoir étendu (spatialement seulement) chaque $\mathcal{V}(x_i, t_i)$ en un ensemble $[\mathcal{V}(x_i, t_i)]_{étendu}$ comprenant 2Npoint, et après avoir décomposé ce dernier en 2 ensembles $[\mathcal{V}(x_i, t_i)]_{R_1}$ et $[\mathcal{V}(x_i, t_i)]_{R_2}$, nous allons pour chaque point de $[\mathcal{V}(x_i, t_i)]_{R_1}$ (resp. $[\mathcal{V}(x_i, t_i)]_{R_2}$) effectuer son estimation $z^*(x_\beta, t_\beta)$ en minimisant la forme quadratique suivante (ajustement par moindre carré) :

$$\sum_{\{x_\alpha, t_\alpha\}} [\ P_{\alpha\, l}(k_{prior}\ ,\ \{f_i\}_{prior}) - z(x_\alpha, t_\alpha)\ ]^2 \tag{3.16}$$

avec :

$$P_{\alpha\, l}(k_{prior}\ ,\ \{f_i\}_{prior}) = \sum_{l=\beta}^{k_{prior}} a_l f_l(x_\alpha) + \sum_{\{f_i\}_{prior}} [a_i \cos(2\pi f_i t_\alpha) + b_i \sin(2\pi f_i t_\alpha)\ ] \tag{3.17}$$

Dans ces expressions, $\{x_\alpha, t_\alpha\}$ est l'ensemble des points de $[\mathcal{V}(x_i, t_i)]_{R_1}$ (resp. $[\mathcal{V}(x_i, t_i)]_{R_2}$), $f_\ell(x_\alpha)$ un polynôme de degré ℓ en x_α et $\{a_l, a_i, b_i\}$ l'ensemble des coefficients à déterminer.

Ces opérations permettent de dégager autant d'erreurs $(z(x_\beta, t_\beta) - z^*(x_\beta, t_\beta))$ qui serviront à l'élaboration d'une variance expérimentale :

$$S_{prior}(ivois) = \sum_{[\mathcal{V}(x_i,t_i)]_{étendu}} \frac{(z(x_a,t_a) - z^*(x_a,t_a))^2}{2Npoint} \qquad (3.18)$$

et il en résultera autant de quantités $S_{prior}(ivois)$ que de zones tests.
Ceci permettra d'établir les moyennes :

$$E(k_{prior}, [f_i]) = \sum_{nvois} \frac{S_{prior}(ivois)}{nvois} \qquad (3.19)$$

C'est par 3.19 que nous quantifions les hypothèses "$k = k_{prior}$" et "$\{f_i\} = \{f_i\}_{prior}$:". En incrémentant k à partir de $\emptyset$ et en faisant prendre aux fréquences f_i les valeurs induites par notre connaissance de la physique du phénomène, nous obtiendrons l'ensemble $[\ k\ ,\ [f_i]\ ]$ qu'on qualifiera d'optimal pour l'opération ultérieure d'estimation.

Inférence de K(h) (boucle-K) :

Une fois $[\ k\ ,\ [f_i]\ ]$ déterminé pour une géométrie $\mathcal{V}$ fixée, la CG K(h) optimale pour ces caractéristiques. $[\mathcal{V}\ ,\ k\ ,\ [f_i]\ ]$ sera déterminée d'une manière classique, puisqu'établie à partir de CLA-k-$\{f_i\}$ qui font passer de l'espace-temps à l'espace. Les modèles classiques de CG sont alors utilisés. Il est à noter que puisque dépendantes seulement d'un incrément spatial h, et puisque plongées dans l'espace-temps, ces CG perdront alors leur stricte positivité conditionnelle.

Validation du modèle (boucle-K*) :

Les étapes précédentes de l'inférence ont fourni plusieurs ensembles $[\mathcal{V}\ ,\ k\ ,\ [f_i]\ ,\ K(h)\ ]$ de paramètres.

Pour déterminer l'ensemble optimal, on va mettre en oeuvre l'opérateur d'estimation caractérisé par cet ensemble de paramètres, et qui est tout simplement le Krigeage en FAI-k-$\{f_i\}$ (cf annexe C-2b). On rappelle que celui-ci possède une composante destinée à la partie périodique et temporelle de z(x,t). A ce détail près, la "boucle-K*" est comparable à celle d'une FAI-k classique : estimations sur les zones tests pour tester les différents ensembles $[\mathcal{V}\ ,\ k\ ,\ [f_i]\ ,\ K(h)\ ]$ et choix de celui offrant le meilleur résultat en moyenne.

3 .1 .5 Voisinages

La VR z(x,t) étant spatio-temporelle, nous aurons à considérer simultanément l'échantillonnage spatial des anomalies Y(x) et l'échantillonnage temporel de D(t). Désignons par $\mathcal{D}^{x\,t}$ et $\mathcal{P}^{x\,t}$ les vecteurs diamètres et pas de $\mathcal{V}$, avec :

$\mathcal{D}^{x\,t} = (\ \mathcal{D}^x\ ,\ \mathcal{D}^t\ )$: vecteur diamètre de $\mathcal{V}$ (3.20)

$\mathcal{P}^{x\,t} = (\ \mathcal{P}^x\ ,\ \mathcal{P}^t\ )$: vecteur pas de $\mathcal{V}$ (3.21)

et :

$(\ \mathcal{D}^x\ ,\ \mathcal{P}^x\ )$: caractéristiques de la section spatiale de $\mathcal{V}$ (3.22)

$(\ \mathcal{D}^t\ ,\ \mathcal{P}^t\ )$: caractéristiques de la section temporelle de $\mathcal{V}$ (3.23)

Nous allons examiner quelles doivent être les caractéristiques optimales de $\mathcal{D}^{x\,t}$ et $\mathcal{P}^{x\,t}$ pour appréhender au mieux chacune des composantes du PTP et voir dans quelle mesure ces caractéristiques idéales sont ou non accessibles.

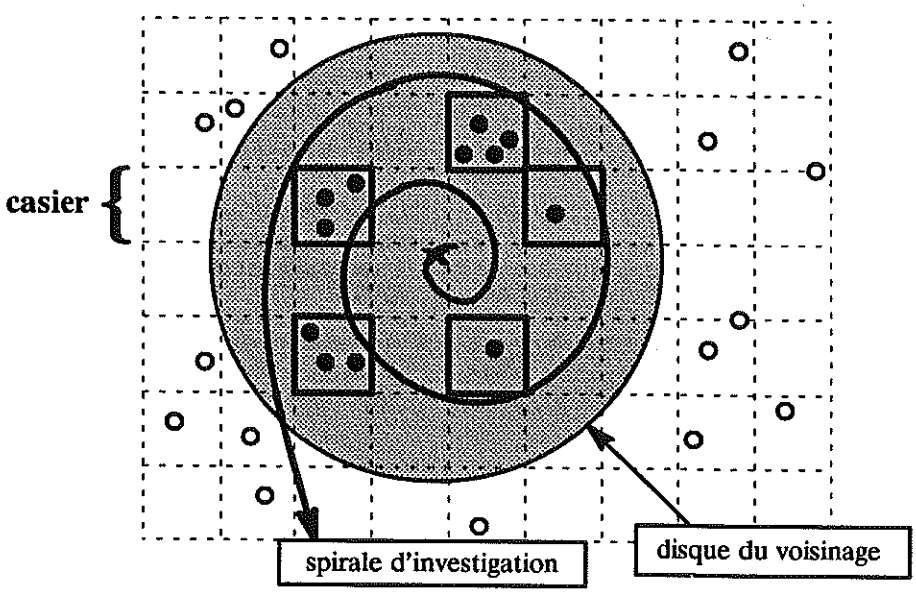
Ces considérations conduiront à proposer 4 configurations géométriques typiques qui seront testées dans les boucles d'inférence. Nous rappellerons au préalable comment la machine procède à la constitution automatique de $\mathcal{V}$.

Constitution automatique du voisinage :

Il faut dire tout d'abord que les programmes à notre disposition, ne "savent" pas oeuvrer dans l'espace-temps (cas de Bluepack, ref [GST-1]) ce qui obligera à pratiquer certaines astuces pour le contraindre à respecter les contraintes idéales d'échantillonnage temporel.
Deux lois régissent cette recherche de voisinage :

1) La proximité spatiale : notons $x = (x_1, x_2)$ le centre de $\mathcal{V}$.
Lorsque la machine aura (pour une direction donnée de l'espace) un choix à effectuer parmi plusieurs échantillons informés, elle prendra celui qui est le plus proche de x.

2) L'isotropie spatiale : si $\mathfrak{D}^x$ représente le diamètre choisi par l'utilisateur et si on trace un disque centré sur x et de diamètre $\mathfrak{D}^x$, la machine effectuera son échantillonnage de sorte que les éléments de $\mathcal{V}$ se répartissent de façon homogène sur ce disque.



: *point ciblé par l'analyse ou l'estimation*

: *autres points informants*

: *points "éligibles"*

: *octant validé car contenant de l'information*

Figure 3.3 : Méthode de recherche automatique de voisinage (logiciel BLUEPACK)

Pour atteindre ces 2 critères, la machine procède comme suit (figure 3.3):

- Par commodité algorithmique, elle décompose la zone d'étude en une partition de "casiers" d'une taille choisie par l'utilisateur et dans lesquels les échantillons seront rangés.

- Elle effectue une investigation par spirale croissante centrée sur x, pour trouver quels casiers contiennent des échantillons.
Ceux qui en contiennent sont dits "validés". La spirale s'arrête lorsque $\mathfrak{D}^x$ est atteint.

- Puis vient le comptage de la population des casiers validés.
Si elle excède le nombre Npoint (nombre de points du voisinage choisi par l'utilisateur), les points pris seront ceux qui seront les plus proches de x.

Cette recherche permet en général d'atteindre l'isotropie de $\mathcal{V}$ (loi n°2 satisfaisante) ainsi que la proximité spatiale (loi n°1).

Elle sera d'autant plus efficace que les données seront uniformément réparties dans le plan.

Section spatiale optimale du voisinage :

Il s'agit de voir quels sont les paramètres ($\mathfrak{D}^x$, $\mathfrak{P}^x$) optimaux pour l'analyse de la FAI-k spatiale Y(x)

En ce qui concerne le diamètre $\mathfrak{D}^x$, cette quantité peut être quelconque, puisque c'est la représentation localement stationnaire (réf **[GST-9]**) qu'on souhaite atteindre lors de l'inférence. En général, $\mathfrak{D}^x$ est choisi de telle sorte que $\mathcal{V}$ contienne une vingtaine de points répartis d'une manière homogène dans le plan.

En ce qui concerne le pas d'échantillonnage $\mathfrak{P}^x$, il peut être arbitraire également, puisque par son biais, on se réfère à l'échantillonnage de la représentation localement stationnaire. Cependant, comme on ne souhaite pas perdre trop du détail de Y(x), ce pas est choisi aussi fin que se peut.

Section temporelle optimale du voisinage :

Il s'agit de voir quels sont les paramètres ($\mathfrak{D}^t$, $\mathfrak{P}^t$) optimaux pour l'analyse de la série trigonométrique D(t) donnée par la formule 3.9. Deux configurations d'échantillonnage sont envisageables :

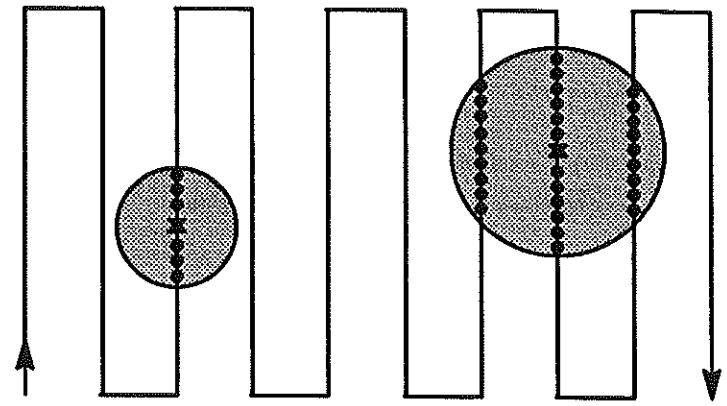
- Un échantillonnage de type "stroboscopique" (ref **[TdS-2]**) pour lequel $\mathfrak{D}^t$ peut être quelconque, pour peu que $\mathfrak{P}^t$ soit supérieur à la plus petite période de la SF 3.9, et que le nombre de points soit suffisant pour pouvoir reconstituer la plus grande période contenue dans la SF 3.9.
- Un échantillonnage tel que $\mathfrak{D}^t$ recouvre au plus la plus grande période contenue dans la SF 3.9. Dans ce cas, les contraintes du théorème d'échantillonnage (ref **[TdS-5]**) imposent, on le rappelle, que $\mathfrak{P}^t$ soit inférieure strictement à la moitié de la plus petite période de la SF 3.9.

Comme l'expérience nous confronte à un échantillonnage quasiment coplanaire il sera difficile d'obtenir exactement l'une ou l'autre de ces configurations. Les géométries proposées se situeront donc à mi-chemin entre ces solutions.

Géométrie du type $\mathcal{V}_1^x$, spatial aveugle :

Si on applique aveuglément l'algorithme de constitution automatique de $\mathcal{V}$ décrite précédemment, le résultat obtenu est celui de la figure 3.4: les points sélectionnés sont alignés sur les profils, et très proches du centre du voisinage.

Cette figure montre deux exemples, selon le rapport existant entre $\mathfrak{D}^x$ et la distance entre les profils. Nous appellerons cette géométrie "voisinage spatial aveugle", et nous la symboliserons par le terme générique $\mathcal{V}_1^x$. L'exposant "x" est là pour rappeler que cette configuration ne tient pas compte de l'échantillonnage temporel résultant.



⊠ : *point ciblé par l'analyse ou l'estimation*

● : *points sélectionnés*

Figure 3.4 : Géométrie du voisinage du type $\mathcal{V}_1^x$ (spatial "aveugle")

La configuration $\mathcal{V}_1^x$ possède les inconvénients suivants:

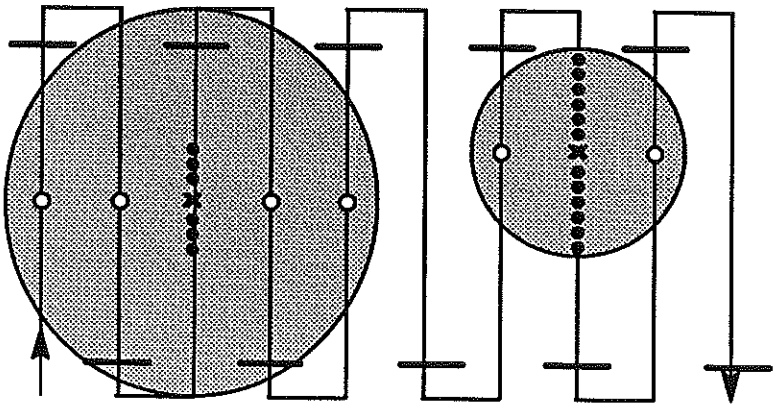
- Elle n'est pas isotrope dans l'espace et donc échantillonne mal l'anomalie spatiale Y(x).
- Comme les points sont très proches spatialement, ils le sont également dans le temps, de sorte que les périodes de D(t) se trouvent également mal échantillonnées.
- On montrera dans le chapitre consacré à l'estimation, que de telles configurations géométriques peuvent rendre impossible l'estimation ultérieure.

Ces problèmes sont uniquement liés au fait que les données sont sur profils. Ils surviennent en toute généralité dès qu'on analyse des données prélevées en mer.

Géométrie du type $\mathcal{V}_2^x$, spatial anisotrope :

L'étape suivante consiste à contraindre $\mathcal{V}$ à s'étendre perpendiculairement aux profils, pour tenter d'atteindre l'isotropie spatiale du voisinage.

Pour ce faire, la route est décomposée en "pseudo-profils" lâches, et la machine impose à $\mathcal{V}$ de reposer sur "Ncode" profils différents.



⊠ : *point ciblé par l'analyse ou l'estimation*

○ : *points contraints de reposer sur des profils différents*

● : *points sélectionnés sans contraintes*

Figure 3.5 : Géométrie du voisinage du type $\mathcal{V}_2^x$ (spatial anisotrope)

La figure 3.5 montre deux exemples, selon le rapport existant entre le nombre de points de $\mathcal{V}$ et le nombre de profils sur lesquels repose $\mathcal{V}$.

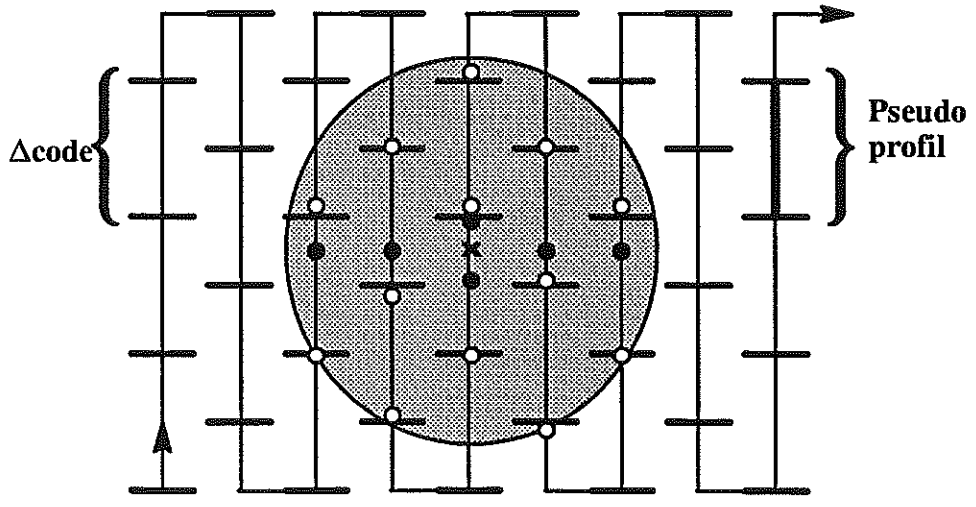
Cette configuration (notée $\mathcal{V}_2^x$) possède les inconvénients suivants:

- Elle n'est toujours pas isotrope dans l'espace,
- L'échantillonnage de D(t) n'est pas nécessairement correct et dépend des caractéristiques de la route.
- Comme dans le cas précédent, de telles configurations peuvent rendre impossible l'estimation ultérieure.
- Ce voisinage est légèrement "sautant" en ce sens que lorsque $\mathcal{V}$ se déplace continûment le long de la route, la machine aura tendance à prendre les points frontières de 2 profils consécutifs. Mais comme ces "pseudo-profils" sont lâches, les points frontières sont rares, et l'effet sautant peu marqué.
  En effet, ces points sont les plus proches du centre du voisinage, tout en reposant sur un autre profil. Il en résulte que la population de $\mathcal{V}$ ne se reproduira pas continûment mais par à-coups.

Nous appellerons cette géométrie "voisinage spatial anisotrope".

Géométrie du type $\mathcal{V}_3^x$, spatial isotrope sautant :

L'étape suivante consiste à contraindre $\mathcal{V}$ à s'étendre sur des "pseudo-profils" de taille plus réduite que précédemment, pour tenter d'atteindre l'isotropie spatiale du voisinage. La figure 3.6 en donne un exemple. Nous appellerons $\mathcal{V}_3^x$ cette géométrie. Si cette configuration permet en général d'atteindre l'isotropie spatiale et de rendre possible l'opération d'estimation, c'est au détriment du pas d'échantillonnage qui devient en moyenne la longueur Δcode des pseudo-profils.



✖ : *point ciblé par l'analyse ou l'estimation*
○ : *points "frontière"*
● : *autres points sélectionnés*

Figure 3.6 : Géométrie du voisinage du type $\mathcal{V}_3^x$ (spatial isotrope "sautant")

Elle possède donc les inconvénients suivants :

- Elle est fortement "sautante" et ceci d'autant plus que Δcode est petit.
- Elle ne réalise pas nécessairement un bon échantillonnage temporel.
- Comme ce sont en majorité les points aux frontières des profils qui sont utilisés, cette géométrie équivaut à remplacer le jeu de données initial par l'ensemble des points frontières.
- Le pas d'échantillonnage devient la longueur Δcode des pseudo-profils. Il en résulte un éventuel recouvrement spectral (réf [TdS-3]) dans le plan si ce pas est trop lâche.

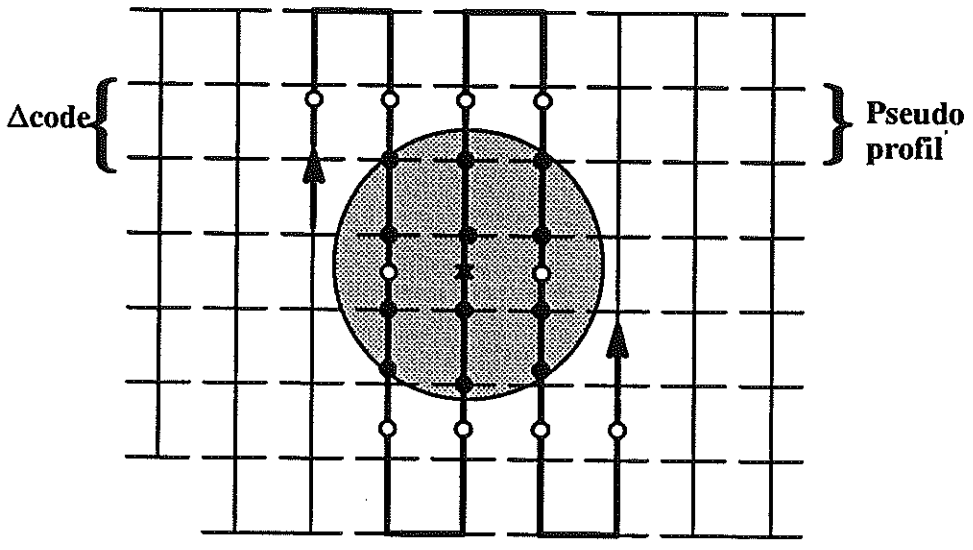
Pour ces raisons, nous appellerons cette géométrie "voisinage spatial isotrope sautant".

Géométrie du type $\mathcal{V}_4^{x\,t}$, spatio-temporel :

La dernière étape consiste à tenter de concilier un voisinage de type $\mathcal{V}_3^x$ avec un meilleur échantillonnage temporel. Dans ce but, on ajoute à la configuration précédente un certain nombre de contraintes portant sur quelques points de $\mathcal{V}$: ceux-ci doivent entourer (dans le temps) le centre du voisinage, être à maille régulière, et glisser dans le temps.

La figure 3.7 montre un exemple de la géométrie résultante. Cette géométrie sera notée $\mathcal{V}_4^{x\,t}$ et appelée "spatio-temporelle". Elle offre l'avantage de bien échantillonner les périodes de D(t), de préserver l'isotropie spatiale de l'échantillonnage et d'atténuer l'effet sautant spatial puisque la maille temporelle est glissante.

Son inconvénient majeur apparaîtra lorsque nous expérimenterons cette géométrie : elle rend l'inférence non robuste.

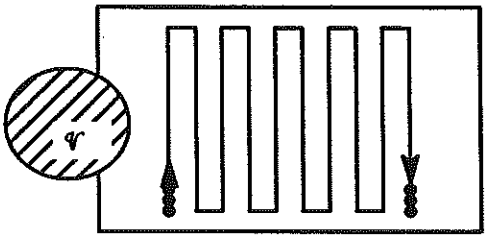
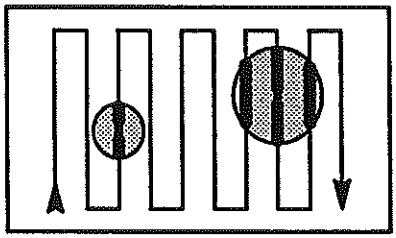
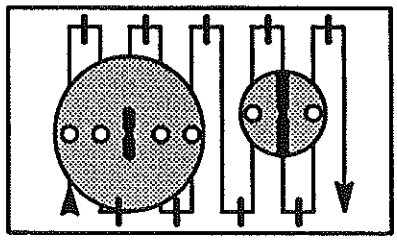
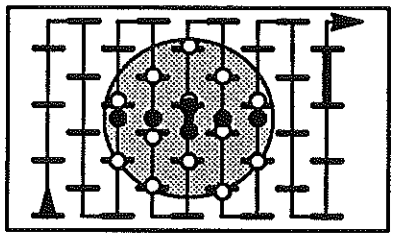
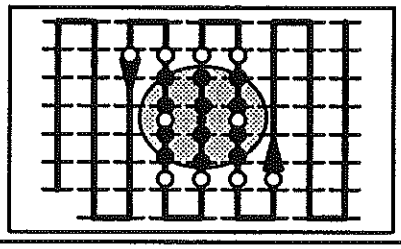


✖ *: point ciblé par l'analyse ou l'estimation*

○ *: points contraints dans le temps*

● *: autres points sélectionnés*

Figure 3.7 : Géométrie du voisinage du type $\mathcal{V}_3^{x\,t}$ (spatio-temporel)

Algorithme de constitution automatique de $\mathcal{V}$	LES PROBLEMES Avantages (+)	Inconvénients (-)
Loi 1 : proximité spatiale **Loi 2 : isotropie spatiale**		⊕ – maille quasi régulière ⊖ – anisotropie totale de l'échantillonnage
Type $\mathcal{V}_1^x$: spatial aveugle **Paramètres : $\mathfrak{D}^x$, Npoint**		⊕ – Simplicité algorithmique ⊖ – Anisotropie spatiale totale ⊖ – Mauvais échantillonnage temporel ⊖ – Instabilités numériques
Type $\mathcal{V}_2^x$: spatial anisotrope "sautant" **Paramètres : $\mathfrak{D}^x$, Npoint, Δcode (grand), Ncode**		⊕ – Simplicité algorithmique ⊖ – Anisotropie spatiale (tendance) ⊖ – Mauvais échantillonnage temporel ⊖ – Instabilités numériques
Type $\mathcal{V}_3^x$: spatial isotrope "sautant" **Paramètres : $\mathfrak{D}^x$, Npoint, Δcode (petit), Ncode**		⊕ – Isotropie spatiale ⊕ – Meilleur échantillonnage temporel ⊖ – Voisinage "sautant" ⊖ – Recouvrement spectral ⊖ – Peu de données sont actives
Type $\mathcal{V}_4^{x,t}$: spatio-temporel **Paramètres : $\mathfrak{D}^x$, $\mathfrak{D}^t$, Npoint, Δcode (petit), Ncode**		⊕ – Quasi isotropie ⊕ – Voisinage glissant dans le temps ⊖ – Voisinage sautant dans l'espace ⊖ – Instabilités numériques

$\mathcal{V}$: voisinage d'analyse — *$\mathfrak{D}^t$: diamètre temporel de $\mathcal{V}$*

$\mathfrak{D}^x$: diamètre spatial de $\mathcal{V}$ — *$\mathfrak{p}^t$: pas d'échantillonnage temporel de $\mathcal{V}$*

Ncode : nombre de pseudo profils différents touchés par $\mathcal{V}$ (Ncode < Npoint)
Δcode : longueur (temporelle) d'un élément du partionnement en "pseudo" profils
Npoint : nombre de points contenus dans $\mathcal{V}$

Figure 3.8 : Genèse d'un voisinage spatio-temporel

Conclusions sur le voisinage :

La figure 3.8 résume l'ensemble des configurations géométriques qui seront utilisées pour l'analyse de ce PTP spatio–temporel. Aucune d'entre elles n'est idéale, du fait des données sur profil et du mauvais échantillonnage de ce champ de séries temporelles. Le voisinage idéal serait bien entendu un cylindre spatio–temporel, mais il est inaccessible dans le cas présent.

L'un des principaux problèmes liés à ces configurations est l'effet sautant de celles–ci. Cet effet, on le rappelle, provient du séquencement en pseudo–profils de la route du bateau , qui est fait une fois pour toutes. Ce séquencement et les configurations résultantes du voisinage, équivalent à un choix particulier de points informants: les points frontières.

Une amélioration sensible devrait donc être possible si ce séquencement glissait avec le voisinage. Mais un tel algorithme nécessiterait une refonte complète du programme d'analyse (Bluepack) et n'a donc pu être développé dans les temps impartis pour la réalisation de cette thèse.

3 .2 Analyse et modélisation des données de la campagne Krakatao

Ce jeu de données va permettre de présenter un premier cas d'étude d'analyse et de modélisation du magnétisme spatio-temporel. Nous nous pencherons plus particulièrement sur les boucles d'inférence de "k", $\{f_i\}$ et K(h). Ces paramètres sont inférés conditionnellement à chacune des géométries $\mathcal{V}_2^x$, $\mathcal{V}_3^x$ et $\mathcal{V}_4^{x\,t}$ présentées en section 3.1. Nous ne donnons ici qu'un résumé de résultats dont le détail est l'objet de l'annexe D.

3 .2 .1 Présentation des données (figures 3.9 à 3.13)

Ces données sont celles d'une campagne réalisée par l'IFREMER (ref [I-3]) en 1985, à l'aide du navire océanographique Jean Charcot. Elles nous ont aimablement été fournies par V. Renard (ref [I-3]) . L'objectif de cette mission était d'étudier la zone de subduction de la fosse de Java et son raccord avec la faille de Sumatra.

La zone d'étude se situe au large du détroit de la Sonde (figure 3.9). Celle-ci couvre une aire d'approximativement 4 degrés carré (figure 3.10). Cette région est l'une des plus actives du globe car elle compte 57 volcans en activité, dont le Krakatao, tristement célèbre par les 20.000 victimes qu'il fit en 1883, lors d'une éruption particulièrement violente.

Une zone test a été sélectionnée, qui couvre un degré carré (111^2 km^2) (figure 3.11). On constate que la route d'échantillonnage est constituée de profils sensiblement parallèles et équidistants. Ces profils sont parallèles à la zone de subduction.

Les tests porteront sur 14.500 échantillons échelonnés toutes les minutes de navigation.

La courbe du magnétisme brut Z(x, t) développé dans le temps (figure 3.12) révèle de fortes périodicités (24^h) dont il est difficile de savoir si elles sont issues des perturbations temporelles ou si elles sont "artificielles" car issues des aller et retours (de longueur 12^h) du bateau . Cette circonstance annule toute possibilité d'une étude temporelle uniquement.

> *Un certain nombre d'expériences furent effectuées dans le sens d'une étude temporelle uniquement de ces données. Celles-ci ne seront pas décrites dans cette thèse, et on consultera (ref [GST-13]) pour en prendre connaissance. On y lira notamment la tentative d'intégrer explicitement ce "parasitage" des aller et retours du bateau, par une modélisation à l'aide de fonctions de dérive du type "toît", ceci dans le but de dissocier D(t) de ces périodes "artificielles" .*
>
> *Les résultats n'ont pas été probants.*

L'examen de la figure 3.12, ainsi que l'analyse du journal de bord de la campagne ont permis de détecter des données aberrantes dans le fichier, pour t variant de $\simeq$ 11 200 à $\simeq$ 13 400 minutes (zone dite "panne-magnéto" sur la figure). Celles-ci ont été supprimées du fichier.

> *Dans l'intervalle de temps précité, on note en effet la présence de valeurs fortement négatives. L'examen du journal de bord du bateau, et notamment le suivi des appareils de mesure, a permis de voir que cela correspondait à une panne du magnétomètre (qui est traîné derrière le bateau), la remontée de celui-ci (à $t \simeq$ 12 000), la réparation, ($t \in$ [12 000 , 13 000]), la remise à l'eau (avec toutes les fluctuations de mesure que cela implique) et la reprise des mesures.*

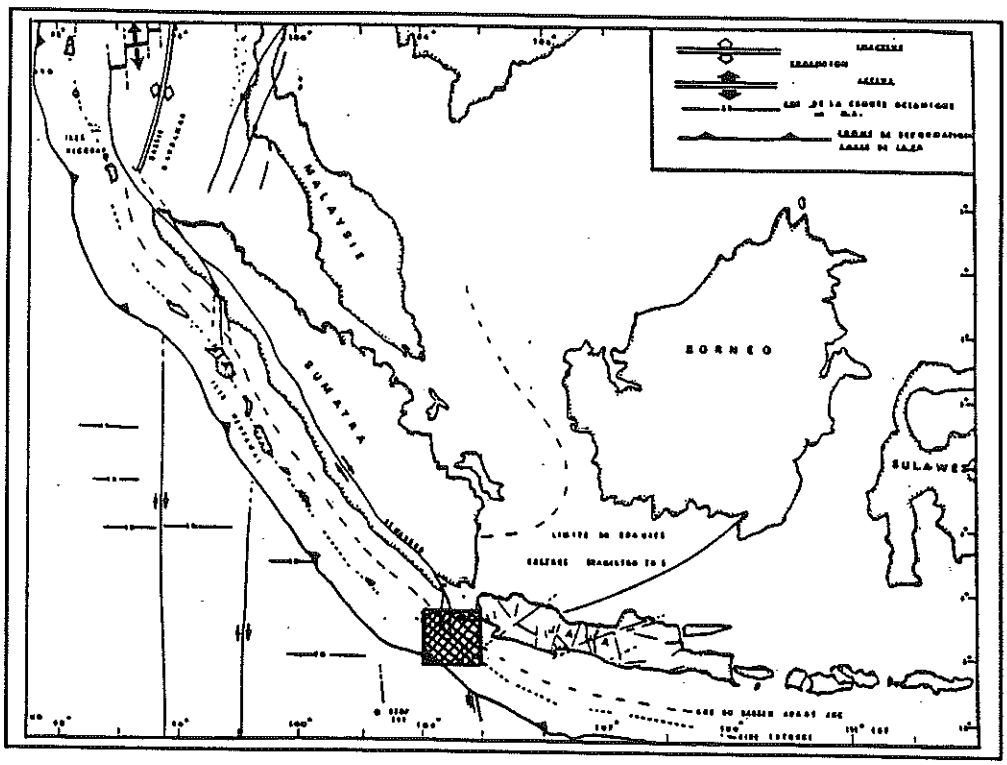


Figure 3.9 : Localisation géographique de la campagne Krakatao – 1985

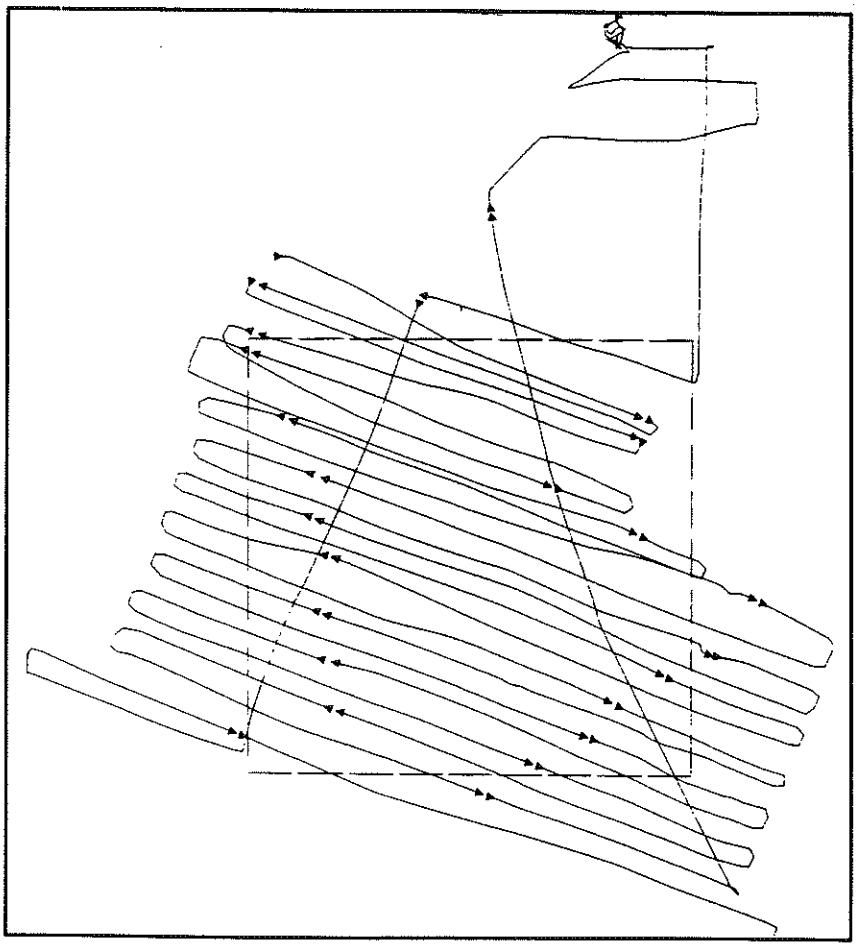


Figure 3.10 : Route du bateau

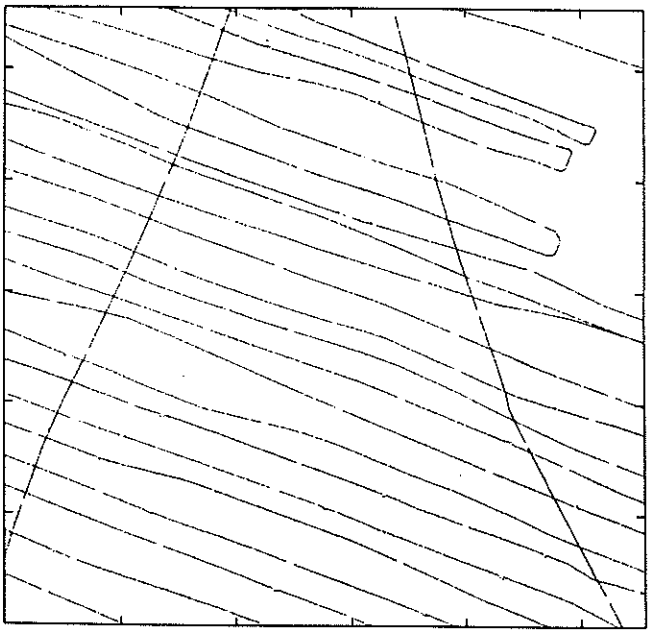


Figure 3.11 : Zone test et profil des expériences

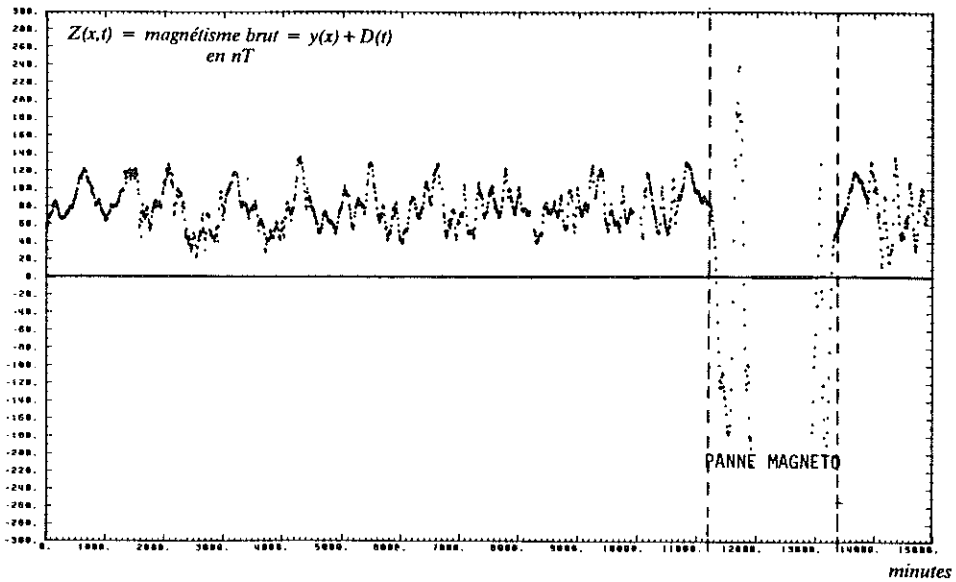


Figure 3.12 : Evolutions dans le temps du magnétisme spatio-temporel z(x,t)

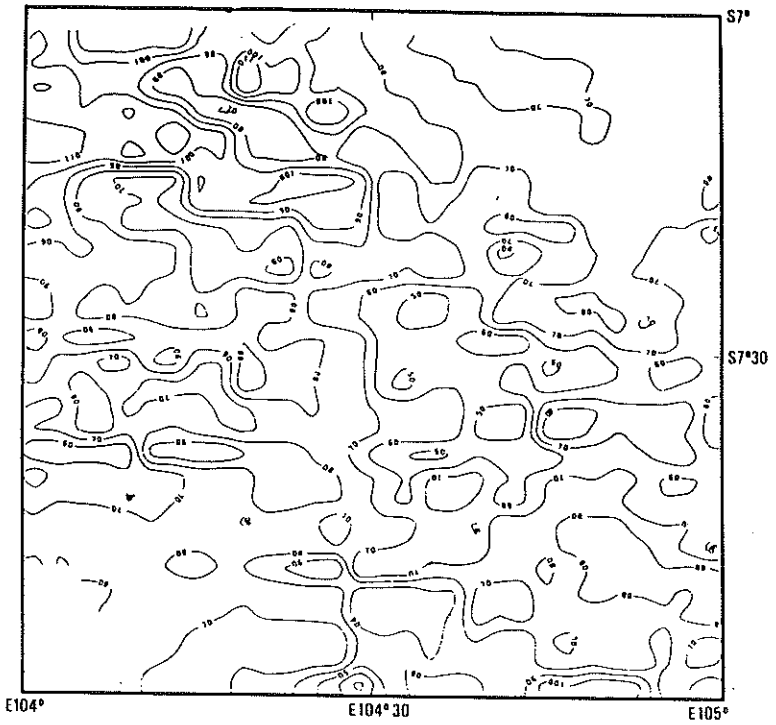


Figure 3.13 : Magnétisme spatio-temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes

La figure 3.13 représente une cartographie (dans le plan) du magnétisme brut z(x, t), sans correction des perturbations temporelles. On y note l'absence de structures spatiales interprétables et l'existence d'un "effet de route" dû aux profils transverses : comme ceux-ci contiennent une information incohérente avec les autres profils qu'ils recoupent, les structures spatiales sont "hachées" par ce passage.
La cartographie résultante ne fait alors que reproduire les profils incohérents. On paye là le tribut de l'absence de filtrage des perturbations diurnes.

3 .2 .2 Les modèles inférés

Les analyses (dont le détail figure en annexe D) ont conduit aux trois modèles suivants :

- $\{ \mathcal{V}_2^x \; ; \; k = 2 \; ; \; T = 24^h \; ; \; K^1(h) \}$

  avec $K^1(h) = a_1 \, |h|$

- $\{ \mathcal{V}_3^x \; ; \; k = 2 \; ; \; T = 24^h \; ; \; K^2(h) \}$

  avec $K^2(h) = a_0 \, \text{pépite}(h) + a_3 \, |h|^3$

- $\{ \mathcal{V}_4^{x\,t} \; ; \; T = 24^h, 12^h \; ; \; k = 2 \; ; \; K^3(h) \}_{\text{prior}}$

  avec $K^3(h) = a_2 \, |h|^2 \, \text{Log}(|h|)$

Dans le troisième modèle, l'indice "prior" existe pour rappeler que ce modèle n'a pas été inféré. Ces trois modèles serviront dans le chapitre 4 pour filtrer D(t) du magnétisme brut z(x,t).

Les faits marquants de ces inférences sont :

- La nécessité d'adopter un compromis entre le critère par rang et celui de variance lors des tentatives de détection des fréquences temporelles.
  Les conflits rencontrés entre ces deux critères illustrent les difficultés de telles détections.

- Les fortes instabilités numériques engendrées par la géométrie de type $\Psi_4^{x\,t}$ qui rendent inutilisable celle-ci. Ces instabilités sont dues au fait que les contraintes temporelles rendent cette géométrie équivalente à une droite de points alignés.

- L'absence de sélectivité de la "boucle-K*" de validation croisée des modèles, et ceci parce que les algorithmes de constitution du voisinage détruisent la géométrie initialement construite.

- La robustesse du paramètre "k" (degré de la FAI), puisque pour diverses géométries différentes de voisinage, on note que c'est toujours la valeur $k=2$ qui a été inférée.

3 .3 Modélisation des données de la campagne Seaperc

Dans ce deuxième cas d'étude spatio-temporel, nous nous penchons dans le détail sur les configurations des échantillonnages spatiaux et temporels résultants des géométries de type $\mathcal{V}_2^x$, $\mathcal{V}_3^x$ et $\mathcal{V}_4^{x\,t}$.

Après avoir présenté les données qui seront analysées, nous proposons 5 modèles différents par les fréquences contenues dans le système d'analyse ainsi que par les configurations $\mathcal{V}$.

Comme bon nombre des paramètres de ces modèles sont arbitraires, cette section ne mérite pas l'appellation "d'analyse", mais seulement de "modélisation". Cet arbitraire sera testé dans le chapitre 4 où ces modèles seront utilisés pour filtrer D(t) du magnétisme total z(x,t). Nous évaluerons alors seulement ce que cet arbitraire a coûté.
Ici encore, nous ne donnons qu'un résumé de résultats dont le détail est l'objet de l'annexe E.

3 .3 .1 Présentation des données (figures 3.14 à 3.18)

Ces données sont celles d'une campagne réalisée par l'IFREMER (ref [I-3]) en 1985, à l'aide du navire océanographique Jean Charcot. Elles nous ont aimablement été fournies par Ph. Huchon (ref [I-2]). Un des objectifs de cette mission était d'étudier une zone de fracture au large du Pérou (figure 3.14) : la zone de fracture de Mendana (ref [GΦ-4]).
Nous nous intéresserons uniquement à la boite "Mendana" qui couvre une zone de 111 km^2 approximativement .

La figure 3.15 montre une route d'échantillonnage du bateau qui est constituée essentiellement de profils parallèles de longueur temporelle 12^h . Ces profils sont ici perpendiculaires à la faille. Les tests porteront sur approximativement 6000 échantillons échelonnés toutes les minutes de navigation.

L'intérêt de ce cas d'étude réside en ceci que nous disposons également des perturbations D(t) enregistrées à l'observatoire de HUANCAYO qui se situe à 600 km de la zone d'étude. On rappelle que cette série temporelle a fait l'objet d'une analyse détaillée dans le chapitre 2. Cette analyse avait révélé la présence de fortes énergies pour les périodes 24^h et 12^h .

En conséquence, il est possible de présenter en figure 3.16 les évolutions comparées des perturbations D(t) et du magnétisme brut z(x,t) développé dans le temps. Cette comparaison permet d'évaluer un rapport $\frac{\text{signal}}{\text{bruit}} = \frac{z(x,\,t)}{D(t)}$ approximativement égal à 3.

On notera la présence de pics fortement négatifs (points 1 ... 4, même figure). Ces valeurs sont dues à une zone de fortes valeurs négatives située au Sud-est de la zone d'étude qui sont périodisées par les allers et retours du bateau. On notera de plus que la détection visuelle de D(t) dans z(x,t) est loin d'être immédiate. Les particularités de cet échantillonnage rendent caduque l'analyse dans le temps seulement de z(x,t).

Les figures 3.17 et 3.18 représentent respectivement les cartographies du magnétisme corrigé par l'enregistrement Huancayo et la cartographie du magnétisme brut sans correction. La comparaison de ces deux cartes permet une fois de plus de constater la nécessité d'un filtrage des perturbations pour une analyse des anomalies spatiales.

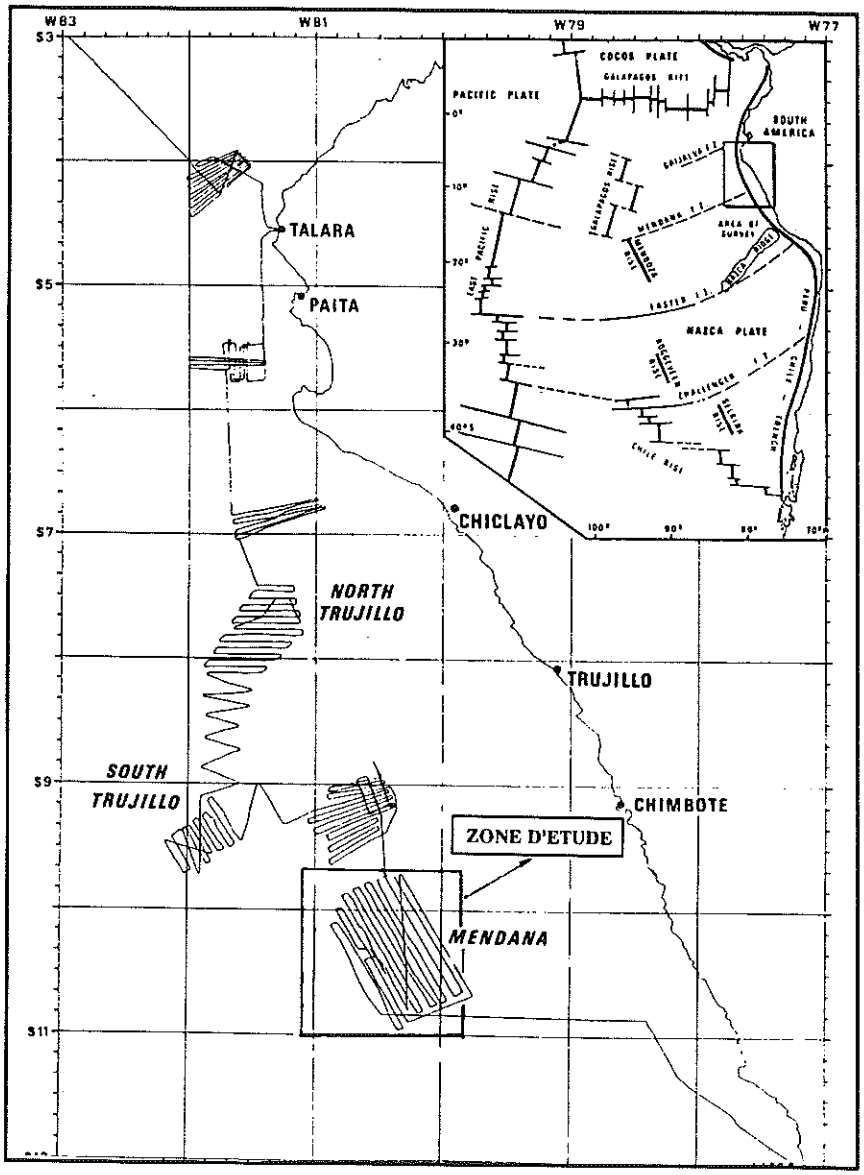


Figure 3.14 : Localisation géographique de la campagne Seaperc – 1985

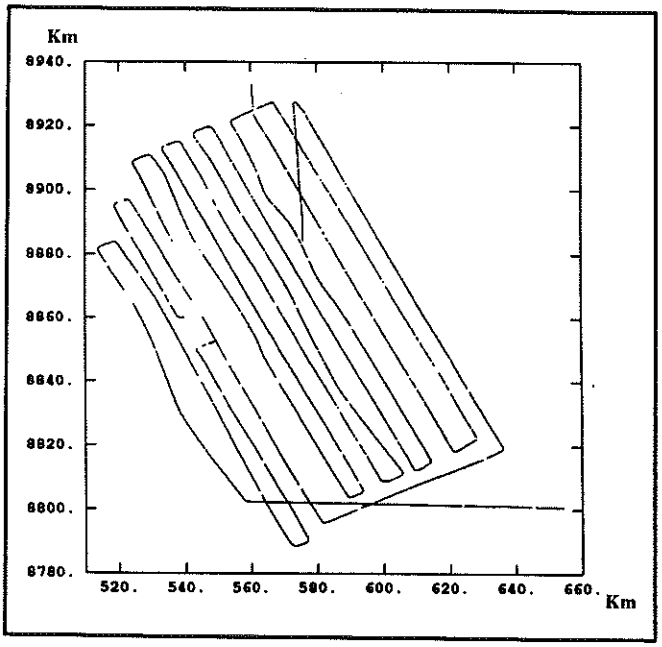


Figure 3.15 : Zone test et profil des expériences

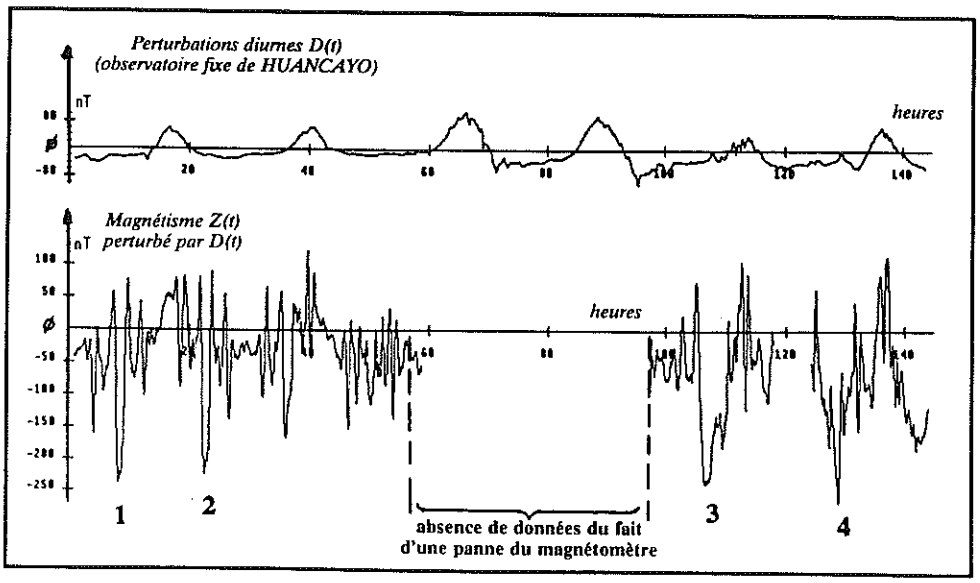


Figure 3.16 : Evolutions dans le temps du magnétisme spatio-temporel z(x,t) comparativement aux perturbations diurnes à Huancayo

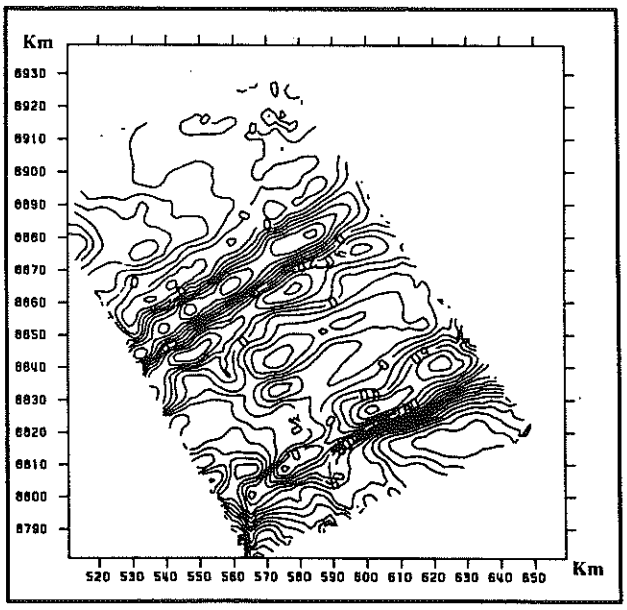


Figure 3.17 : Cartographie des anomalies spatiales y(x) = z(x,t) – D(t) où D(t) est la fonction perturbations diurnes à Huancayo

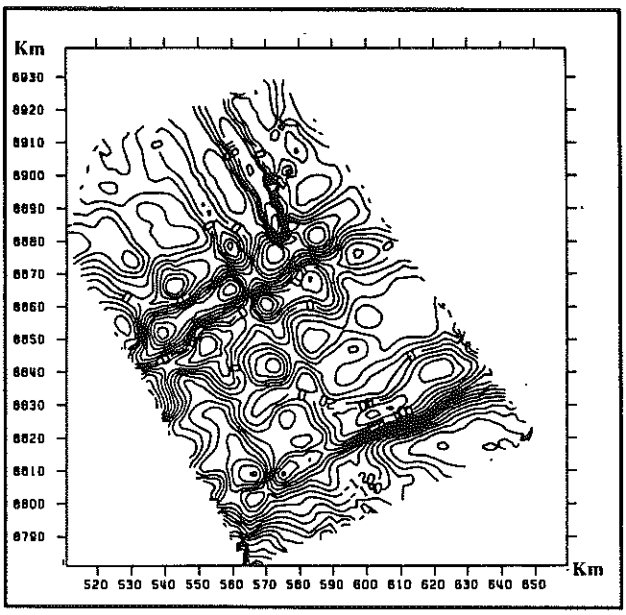


Figure 3.18 : Magnétisme spatio-temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes

3 .3 .2 Les modèles inférés

Les analyses ont permis d'aboutir à 5 ensembles de paramètres dont certains arbitraires sont indexés par "prior" :

- $\{ \gamma_2^x ; k = 1 ; T_{prior} = 24^h ; K^1(h) \}$

  avec $K^1(h) = a_2 \, h^2 \, Log(|h|)$

- $\{ \gamma_2^x ; k = 1 ; T_{prior} = 24^h, 12^h ; K^2(h) \}$

  avec $K^2(h) = a_2 \, h^2 \, Log(|h|)$

- $\{ \gamma_3^x ; k = 1 ; T = 24^h, 12^h ; K^3(h) \}$

  avec $K^3(h) = a_\emptyset$ pépite$(h) + a_2 \, h^2 \, Log(|h|)$

- $\{ \gamma_4^{x\,t} ; k_{prior} = 1 ; T_{prior} = 24^h, 12^h ; K^{prior}(h) = K^3 \}_{prior}$

  avec $K^4(h) = K^{prior}(h) = K^3(h)$

- $\{ \gamma_4^{x\,t} ; k_{prior} = \emptyset ; T_{prior} = 24^h, 12^h ; K^5(h) \}_{prior}$

  avec $K^5(h) = a_1 \, |h|$

On trouvera en annexe E quelques compléments sur l'inférence des différents paramètres du modèle, et surtout un examen minutieux des différentes géométries résultantes des contraintes de voisinage. Les faits marquants de ces inférences sont :

- Une confirmation de la robustesse du paramètre "k" précédemment pressentie, puisqu'ici encore, c'est toujours la même valeur qui est utilisée (k = 1) et ceci pour des géométries très différentes.

- Une confirmation de la nécessité d'abandonner la géométrie $\gamma_4^{x\,t}$.

- Seule la géométrie γ_3^x est acceptable pour appréhender des fréquences temporelles dans ce contexte spatio-temporel particulièrement mal échantillonné.

- Un fait nouveau est apparu : celui que pour 3 modèles, la covariance généralisée est sensiblement la même, et ceci malgré des géométries différentes.
  Pour ce jeu de données, le paramètre K(h) est donc robuste.

3 .4 Analyse et modélisation des données de la campagne Mohns

Ce dernier cas d'étude spatio-temporel offre 4 particularités :

- Nous disposons de 3 jours d'enregistrement des perturbations sur la zone d'étude. Cette série temporelle avait fait l'objet d'une analyse dans le chapitre 2. On avait constaté alors qu'aucune période n'émergeait du spectre.
- Cette campagne est située près du pôle nord. Or on sait que sous cette latitude, les perturbations diurnes sont d'énergie faible par rapport aux anomalies spatiales.
- Les anomalies spatiales sont très hétérogènes, de sorte que nous conduirons une analyse globale (toute la zone d'étude) et une analyse locale (le centre de la zone seulement).
- Dans le but d'obtenir une "boucle-K*" sélective, une présélection arbitraire de 1 donnée sur 10 sera faite.

L'intérêt de ce cas d'étude est donc d'une part de montrer ce qu'il advient lorsqu'on cherche des périodes temporelles dans une VR qui n'en contient pas, et d'autre part de permettre une comparaison entre approche globale et approche locale.
Ici encore, nous ne donnons qu'un résumé de résultats dont le détail est l'objet de l'annexe F

3 .4 .1 Présentation des données (figures 3.19 à 3.23)

Ces données sont celles d'une campagne réalisée par l'IFREMER (ref **[I-3]**) à l'aide du navire océanographique Jean Charcot. Elles ont été mesurées du 4 au 27 juillet 1988, en vue de l'investigation de la dorsale de MOHNS.
Cette campagne revêt pour nous un caractère plus personnel que les autres campagnes, puisque P. Chauvet (ref **[I-1]**) et moi-même y avions participé, sous la direction de V. Renard (ref **[I-3]**) qui nous a aimablement fourni ces données.

La zone d'étude se situe au Nord-Est de l'île de Jean Mayen (longitude 72°N, latitude 1°E, figure 3.19). Celle-ci couvre une aire d'approximativement 240 x 210 km^2. La route d'échantillonnage est ici encore constituée de profils dont certains sont parallèles et d'autres orthogonaux formant ainsi une "croix" contenant de nombreux points de recoupement (figure 3.20).

> *Cette configuration de la route du bateau est, de par sa forme, unique dans les annales de l'IFREMER. Elles est issue des réflexions des scientifiques (dont nous faisions partie) lors de la zone de transit qui nous conduisait de Brest à Jean Mayen. V. Renard nous posait alors la question suivante : "Nous disposons de 11 jours sur zone, pour étudier une dorsale de plusieurs milliers de kilomètres de long, dont l'ouverture est de plus d'une centaine de kilomètres. On souhaite connaître du détail de la ride, mais également de l'information latérale. Quelle doit être la route du bateau, sachant qu'en plus, on souhaite tester l'influence des densités d'échantillonnage sur certains outils géostatistiques?".*
> *Le résultat de nos cogitations collectives fut cette route. ...*

On constate des densités variables de profils, selon que l'on se trouve sur la fracture (zone centrale) ou aux extrémités de la croix. Ceci explique que deux zones seront analysées :

- Une première zone, dite "globale", qui contient l'intégralité de la route du navire . Cette zone est représentée par un polygone sur la figure 3.20.
- Une seconde zone, dite "locale" qui ne contient que la zone centrale de la figure, celle où la densité des profils est particulièrement importante.
  Cette zone est représentée par un polygone sur la figure 3.21.

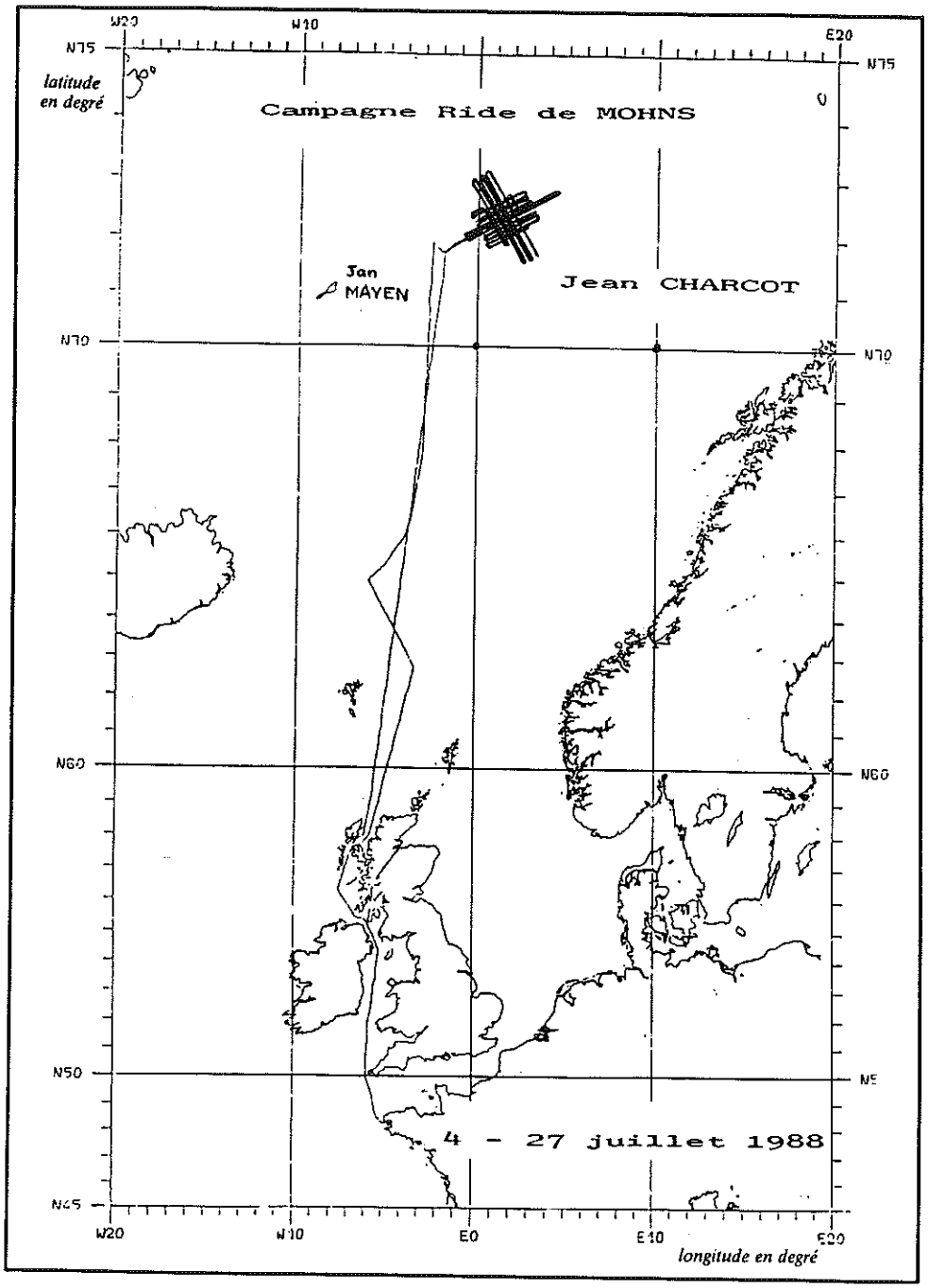


Figure 3.19 : Localisation géographique de la campagne Mohns – 1988

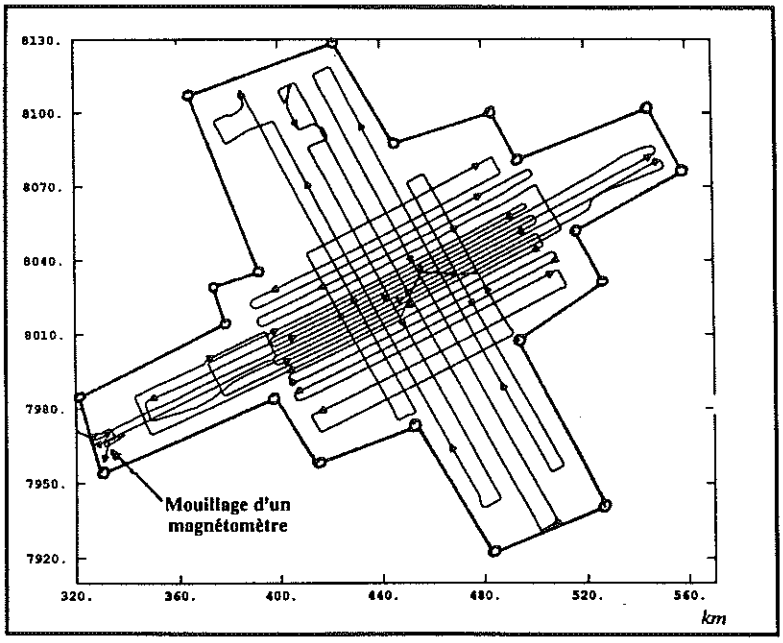


Figure 3.20 : Zone test et profil des expériences (zone dite "globale")

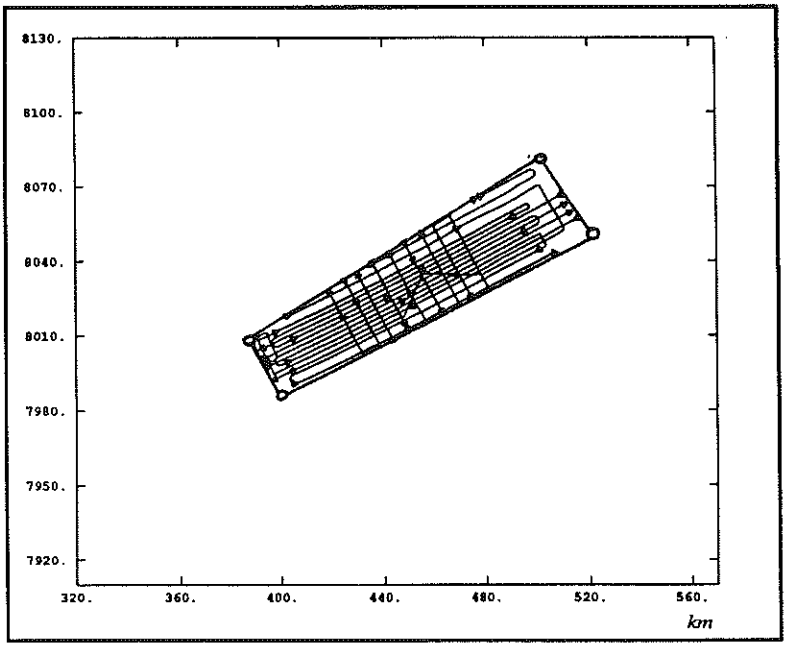


Figure 3.21 : Zone dite "locale"

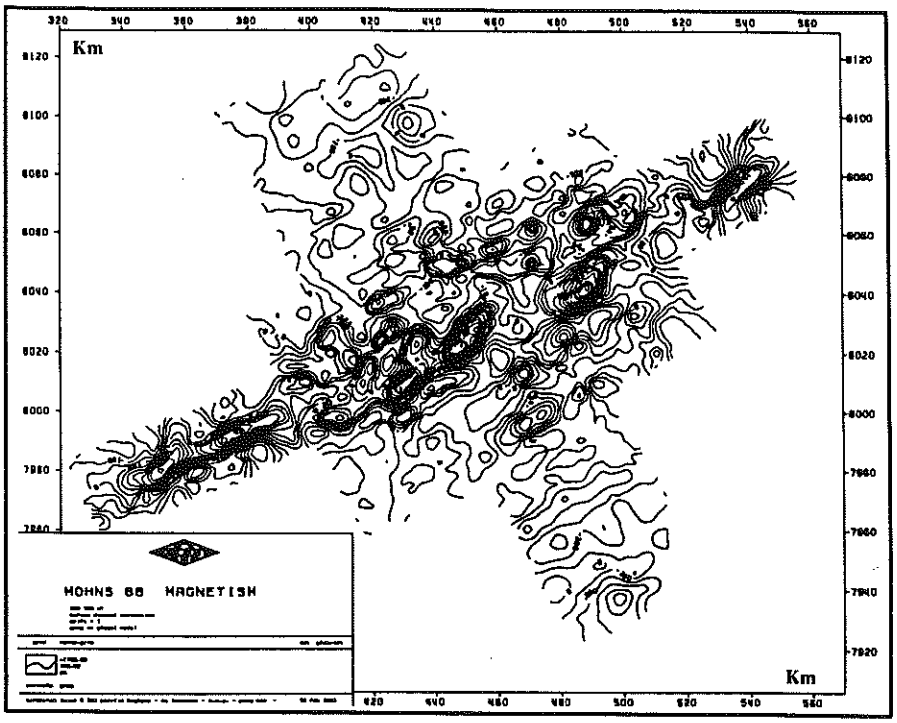


Figure 3.22 : Magnétisme spatio-temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes

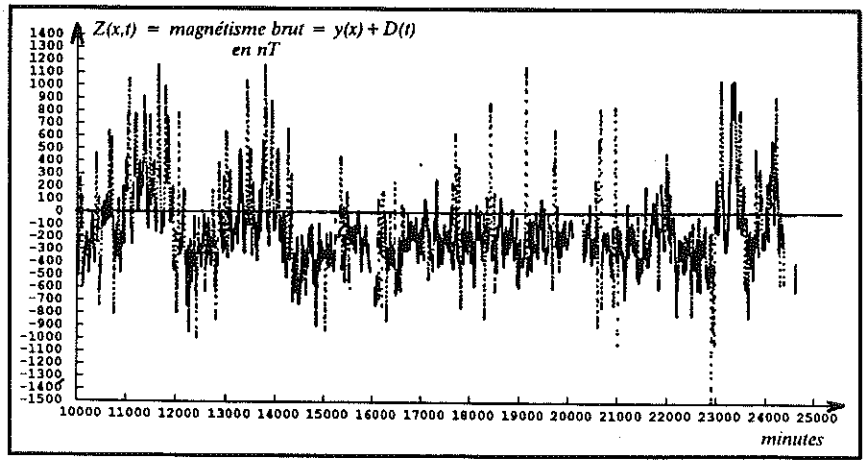


Figure 3.23 : Evolutions dans le temps du magnétisme spatio-temporel z(x,t)

Chacune de ces zones fait l'objet d'analyses et de modélisations par voisinages glissants avec des géométries du type V_2^x ou V_3^x. Les tests porteront sur approximativement 15400 échantillons pour la zone globale et 6500 échantillons pour la zone locale. Les points informants sont échelonnés toutes les minutes de navigation.

La figure 3.22 montre une cartographie dans l'espace du magnétisme total z(x,t) non corrigé des perturbations diurnes. On y constate que la zone locale est plus variable que les extrémités, certes moins informées. Ceci se comprend aisément lorsqu'on sait que la campagne est centrée exactement sur la dorsale de Mohns qui en cet endroit possède une orientation SO-NE. Comme la dorsale est le lieu d'une importante activité volcanique, il est possible de distinguer plusieurs volcans sur cette figure (les structures à fort gradient). En conséquence, la zone locale diffère de l'ensemble de la zone non seulement par une densité plus grande de profil, mais également par une variabilité du magnétisme nettement supérieure. L'adoption de cette zone devrait donc permettre d'isoler du reste ces deux facteurs :

- Hétérogénéité de l'échantillonnage.
- Hétérogénéité de Z(x ,t).

La courbe du magnétisme brut Z(x, t), développée dans le temps (figure 3.23) ne révèle pas la période 24^h. La seule périodicité décelable concerne les 5000 premières minutes du champ (2 périodes de 2500 minutes). Elle est due à la zone centrale (à fort gradient) recoupée deux fois par le bateau.

On note que Z(x,t) présente une forte variabilité qui étend son domaine à l'intervalle [– 1 400 nT, 1 200 nT]. Si on compare ces variations à celles données par les 3 jours de mesure des perturbations qui avaient révélé alors une amplitude crête à crête de 450 nT, on constate que le rapport $\frac{\text{signal}}{\text{bruit}} = \frac{z(x,t)}{D(t)}$ (calculé sur le même champ temporel) est approximativement égal à 5.8. Ce rapport est important comparativement à celui rencontré dans Seaperc. Il illustre parfaitement la description qui a été faite en section 3.1 des perturbations diurnes en magnétisme.

3 .4 .2 Les modèles inférés

Les analyses ont permis d'aboutir à 4 ensemble de paramètres:

- $\left\{ \mathcal{V}_2^x \; ; \; k = 1 \; ; \; T = 24^h \; ; \; K^1(h) \right\}^{\text{global}}_{\text{prior}}$

  avec $K^1(h) = a_1 \, |h|$

- $\left\{ \mathcal{V}_2^x \; ; \; k = 2 \; ; \; T_{prior} = 24^h \; ; \; K^2(h) = K^{prior}(h) \right\}^{\text{local}}$

  avec $K^2(h) = a_\beta \, \text{pépite}(h) + a_3 \, |h|^3$

- $\left\{ \mathcal{V}_3^x \; ; \; k = 1 \; ; \; T = 24^h \; ; \; K^3(h) = K^{prior}(h) \right\}^{\text{global}}_{\text{prior}}$

  avec $K^3(h) = a_1 \, |h|$

- $\left\{ \mathcal{V}_3^x \; ; \; k = 2 \; ; \; T_{prior} = 24^h, 12^h \; ; \; K^4(h) \right\}^{\text{local}}$

  avec $K^4(h) = a_1 \, |h|$

On trouvera en annexe F des compléments sur la manière dont ces paramètres ont été obtenus. Les faits marquants de ces inférences sont :

- L'absence d'ondulations temporelles dans ce jeu de données se manifeste lors de nos inférences, par des incohérences totales entre les différents tests et d'importantes instabilités numériques.
  Ces faits montrent que comme MOHNS n'est pas un PTP, il n'est pas opportun d'associer les perturbations temporelles à une série de Fourier. En d'autres termes, ce ne sont pas ici des CLA-k-$\{f_i\}$ qu'il faut manipuler.

- Malgré le peu de crédit à accorder à ces tests, on peut cependant percevoir que le passage du global au local permet de diminuer les incohérences entre tests et d'obtenir de meilleurs résultats en termes de variances expérimentales.

- Comme il n'y a pas de périodes temporelles, et comme nous sommes capable ici de calculer le variogramme $\sigma_D(t)$ des perturbations qui ont été enregistrées pendant 3 jours (ce qui est peu mais tout juste suffisant pour ce calcul), il paraît envisageable d'ici utiliser des modèles de covariance répondant à la dichotomie :

$$K(h, t) = K_Y(h) + \sigma_D(t)$$

3 .5 Conclusions du chapitre 3

La principale leçon à tirer de ces analyses, c'est la difficulté de détecter des fréquences temporelles, et plus exactement d'obtenir, par nos tests statistiques, une confirmation des ondulations temporelles qui sont pressenties.
Cette circonstance est très grave, car si on est (comme pour Krakatao) dans un cas où nous ne disposons d'aucune information sur le comportement des perturbations, on ne peut se fier d'une manière aveugle à nos techniques de détection. Il faut agir avec beaucoup de minutie, examiner (comme décrit en annexes D et suivantes) dans le détail les résultats pour pouvoir s'orienter vers le modèle le plus adapté aux données.

La raison principale de ces difficultés ne tient pas tant dans le choix d'assimiler les perturbations à une série de Fourier tronquée, ni encore d'assimiler l'invariant spatial à la réalisation d'une FAI-k, non plus que dans les principes méthodologiques qui régissent nos boucles d'inférence. Le fond du problème se situe dans la configuration des données qui sont alignées sur des profils. En effet, ces configurations font que notre recherche automatique de voisinage aura tendance à engendrer des voisinages constitués d'une ou deux droites de points, ce qui d'une part nuit à l'inférence d'une FAI pour k supérieur à 1, et d'autre part engendre un échantillonnage temporel particulièrement désastreux.
C'est la volonté d'échapper à cette "tendance" à l'alignement qui a permis d'aboutir à une géométrie de type $\mathcal{V}_3^*$. Celle ci s'est avérée la plus adaptée à ce problème, et ceci malgré le choix arbitraire des points frontières qu'elle implique. Si cette configuration peut être qualifiée de raisonnablement isotrope (dans le plan), et si elle permet en général (mais sans certitude) d'échantillonner correctement la tendance temporelle, elle possède un inconvénient non négligeable : elle est "sautante" en ce sens (on le rappelle) que lorsque la fenêtre d'analyse (et d'estimation) se déplace continûment dans l'espace-temps, la population du voisinage ne se reproduit pas continûment mais par sauts. Il en résulte un estimateur qui possédera de nombreuses hautes fréquences pouvant être qualifiées de "parasitaires".
Pour remédier à ce dernier mal, les suggestions suivantes devraient permettre une nette amélioration pour l'analyse de ces PTP, et plus généralement l'analyse (et l'estimation) de données sur profil :

- Pratiquer un séquencement en pseudo-profils qui soit glissant dans l'espace.

- Renoncer à l'extension des voisinages et à leur dichotomie en 2 ensembles, lors des boucles d'inférence.

- Lorsqu'un point est retiré de la géométrie pour le réestimer, ne pas compléter l'ensemble par un point proche de la cible, mais par un point respectant la géométrie imposée. En d'autres termes, si Npoint est demandé, constituer $\mathcal{V}$ sous contraintes géométriques à l'aide de Npoint+1 éléments.

Ces modifications sont simples à énoncer, mais difficiles à réaliser informatiquement car elles nécessitent de reconsidérer intégralement le programme d'analyse utilisé. Ceci explique que nous n'ayons pas pu les réaliser dans les temps impartis pour cette thèse.

Si on se place par la pensée dans le cas favorable où ces modifications sont réalisées, on devrait voir disparaître un certain nombre de problèmes, mais on se heurtera toujours à celui, incontournable de la déjà évoquée "coplanarité" de l'échantillonnage spatio-temporel du magnétisme. Cet échantillonnage signifie qu'on ne peut proposer qu'une surface de données (dans l'espace-temps) à un modèle qui nécessite un volume pour pouvoir s'exprimer pleinement. Ce fait signifie qu'on ne pourra jamais éviter qu'au fil de l'analyse ou de l'estimation et puisque l'espace et le temps sont liés par une vitesse, on rencontre périodiquement des configurations spatiales qui engendrent des configurations temporelles qui soient critiques et déstabilisent nos tests basés sur des variances expérimentales.

Il existe cependant un fait expérimental qui devrait atténuer le caractère quelque peu pessimiste de ces conclusions : le cas d'étude Mohns a montré que si les périodes temporelles n'étaient pas présentes dans le signal, il s'avérait impossible de les inférer. Rétrospectivement, ce fait donne quelque valeur aux modèles associés à Krakatao et Seaperc qui eux ont permis de "pressentir un frémissement" de période temporelle...

CHAPITRE 4

ESTIMATION DE PTP SPATIO-TEMPORELS (MAGNÉTISME)

RÉSUMÉ

Dans ce chapitre, nous expérimentons un système de Krigeage Trigonométrique dans le but d'estimer les perturbations diurnes à partir des données spatio–temporelle de magnétisme brut z(x,t) .

Ces estimations portent sur les 3 jeux de données Krakatao, Seaperc et Mohns qui ont fait l'objet d'une analyse dans le chapitre 3 .

Les modèles utilisés pour conduire l'estimation, sont ceux inférés lors de ces analyses .

Les principaux résultats de ce chapitre sont :

- Pour Krakatao : résultats probants malgré l'absence de vérifications expérimentales.
- Pour Seaperc : bons résultats, confirmés par les vérifications expérimentales .
- Pour Mohns : filtrage impossible par krigeage trigonométrique, car Mohns n'est pas un PTP .

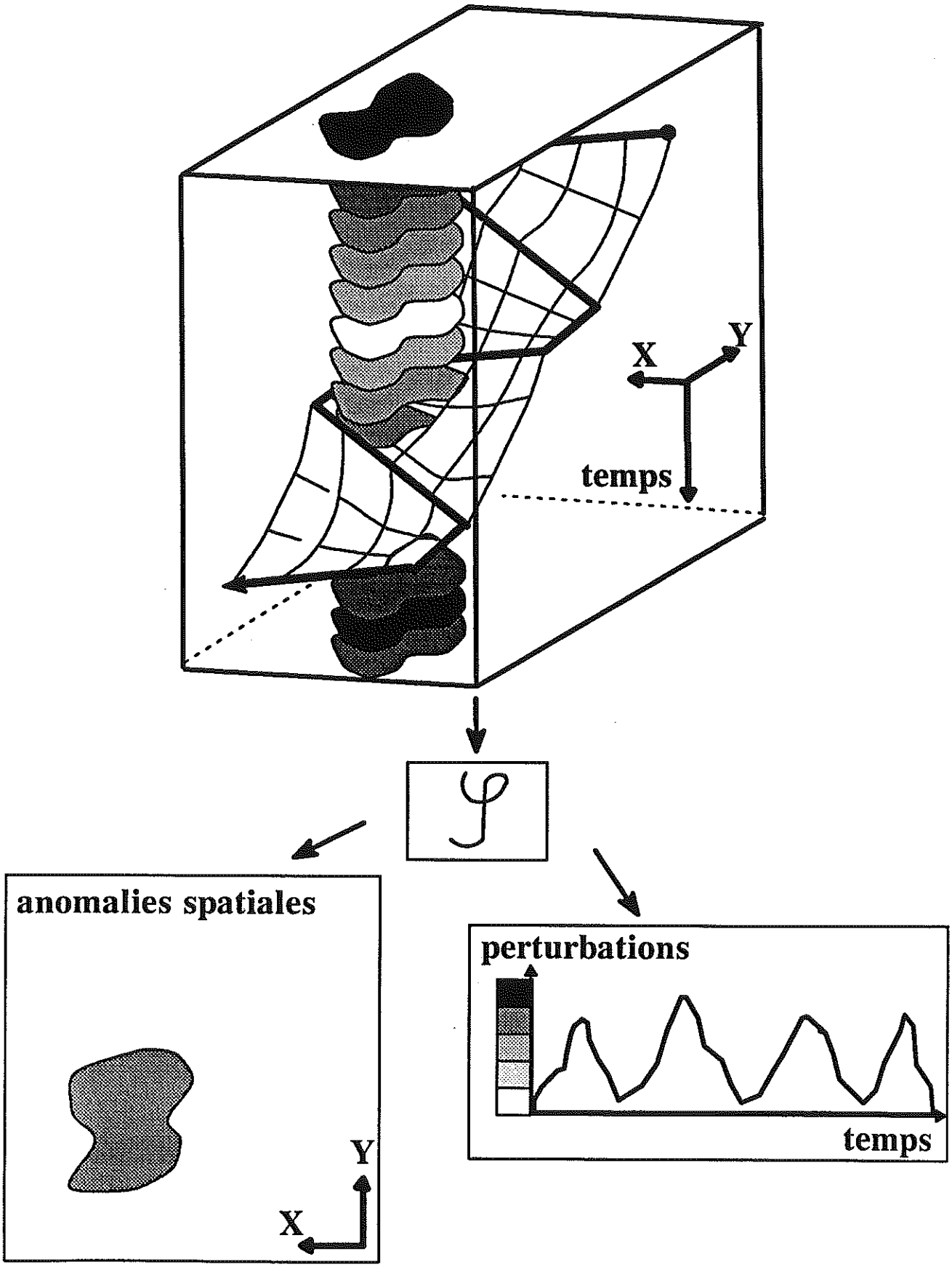
4.1 Contexte des expérimentations . page 99

4.2 Filtrage des données de la campagne KRAKATAO page 101

4.3 Filtrage des données de la campagne SEAPERC page 113

4.4 Filtrage des données de la campagne MOHNS page 132

4.5 Conclusions du chapitre 4 . page 148


X
Y
temps
anomalies spatiales
Y
X
perturbations
temps

Chapitre 4

ESTIMATION DE PTP SPATIO-TEMPORELS (MAGNÉTISME)

Dans ce chapitre, nous tentons d'extraire la composante périodique D(t) (liée aux perturbations diurnes) du magnétisme spatio-temporel z(x,t). Le système utilisé est celui d'un Krigeage Trigonométrique (KT) introduit en annexe B et noté $[\mathcal{S}]$ dans la suite .

Les modèles utilisés pour construire la matrice de $[\mathcal{S}]$, seront ceux inférés dans le chapitre 3 sur les jeux de données Krakatao, Seaperc et Mohns .

On rappelle que pour chaque jeu de données, plusieurs modèles sont disponibles . Ils diffèrent par :

- la géométrie de $\mathcal{V}$ qui leur est associée ,
- les paramètres "k" (degré de la FAI) et K(h) (CG) qui caractérisent la FAI-k spatiale contenue dans z(x,t) et que nous noterons Y(x) ,
- les fréquences $\{ f_i \}$ injectées dans $[\mathcal{S}]$ sous la forme de fonctions de dérive trigonométriques temporelles dans le but de modéliser l'autre composante de z(x,t) : les perturbations D(t)
- "l'a-priori" de certains paramètres qui ont dû être imposés au modèle et non inférés.

Ce qui différencie chacun des cas d'étude, c'est essentiellement que :

- Pour Krakatao, les anomalies spatiales Y(x) sont très fluctuantes .
- Pour Seaperc, D(t) est fortement périodique .
- Pour Mohns, la perturbation D(t) n'est pas un PTP .
- Pour Mohns encore, des estimations globales (le champ complet) et locales (modèle inféré sur une partie du champ seulement) sont conduites .

Certaines vérifications seront possibles, puisque pour Seaperc et Mohns, nous disposons d'enregistrements (séries Huancayo et Mohns) qui constituent une approximation acceptable de la réalité des perturbations sur la zone d'étude .

4 .1 Contexte des expérimentations

Les estimations qui suivent seront essentiellement conduites en voisinage glissant .

4 .1 .1 Les géométries du voisinage

On rappelle que les géométries pour $\mathcal{V}$ sont de 3 types :

- $\mathcal{V}_2^x$: spatial anisotrope
  On rappelle que l'indice supérieur "x" signifie que les contraintes géométriques portent seulement sur l'espace.
  Cette géométrie est en général constituée d'un amas de points alignés sur le profil porteur du point ciblé par l'estimation, et de quelques autres points prélevés sur d'autres profils.
  Elle n'est pas isotrope dans l'espace .
- $\mathcal{V}_3^x$: spatial isotrope sautant .
  Les points sont tous répartis sur des "pseudo-profils" courts, de sorte que la géométrie

approche l'isotropie d'une manière acceptable .
Mais elle offre l'inconvénient d'être fortement "sautante" puisque lorsque $\mathcal{V}$ glisse le long du profil porteur, la population du voisinage ne se renouvelle pas continûment mais par sauts .

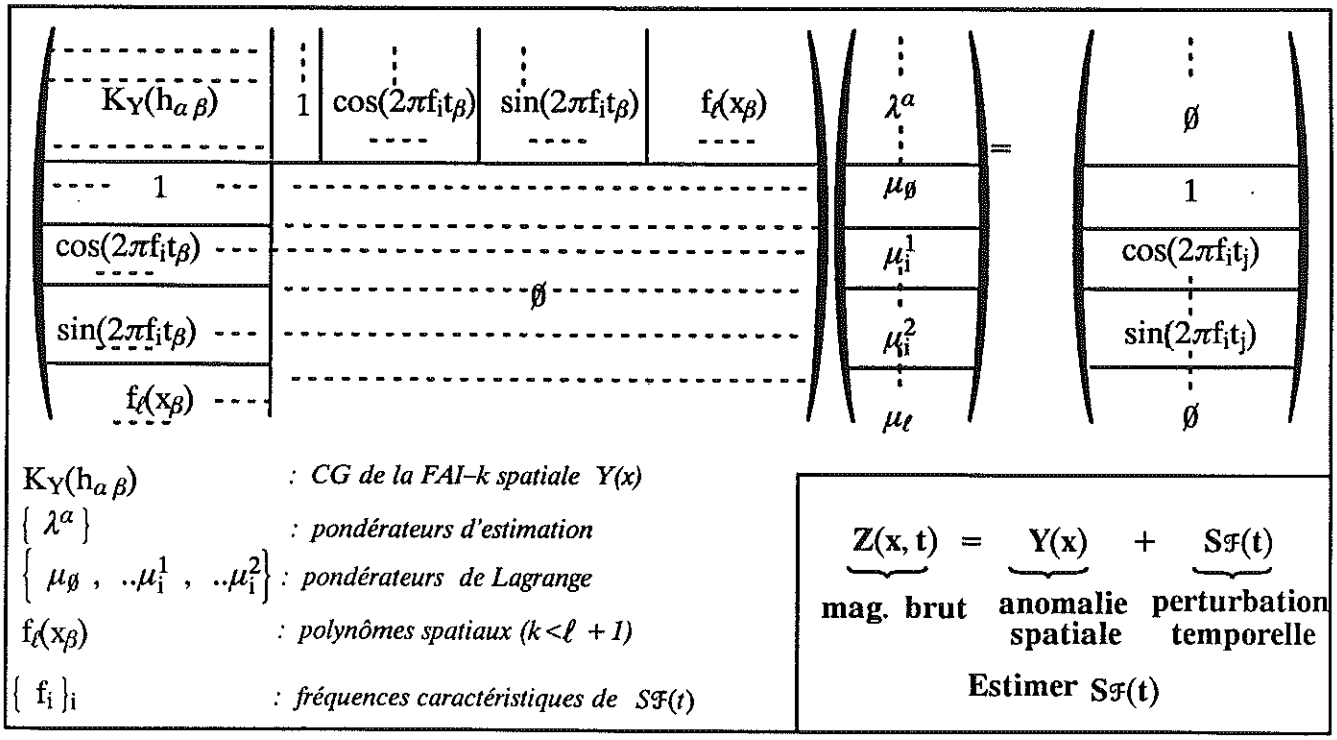
- $\mathcal{V}_4^{x\,t}$: spatio–temporelle .
  Cette géométrie est constituée de la réunion d'un voisinage de type $\mathcal{V}_3^x$ et de points contraints à encadrer dans le temps le temps ciblé par l'estimation .

On rappelle les propriétés de ces géométries :

- le Krigeage avec $\mathcal{V} = \mathcal{V}_2^x$ est instable car des points sensiblement alignés ne permettent pas l'inférence correcte (ni l'inversion de $[\mathcal{S}]$) pour $k > \emptyset$.
- le Krigeage avec $\mathcal{V} = \mathcal{V}_3^x$ donne les meilleurs résultats pour ces données sur profils, mais son caractère sautant risque de bruiter l'estimation .
- le Krigeage avec $\mathcal{V} = V_4^{x\,t}$ est instable à cause du contraste d'échelle entre les points contraints temporellement et spatialement .
  Ce contraste rend $\mathcal{V}_4^{x\,t}$ équivalent à une géométrie du type $\mathcal{V}_2^x$.

4 .1 .2 Système linéaire utilisé pour l'estimation

C'est un système de Krigeage Trigonométrique (KT) pour lequel donc le second membre est nul partout sauf dans les cases liées aux fréquences d'intérêt .
On trouvera en annexe B son élaboration heuristique.
La forme matricielle de ce système noté $[\mathcal{S}]$ est :



4 .2 Filtrage des données de la campagne KRAKATAO

Cette section se caractérise par l'absence de vérifications expérimentales . C'est pourquoi il sera fait appel à un géophysicien pour juger de la qualité des résultats .

4 .2 .1 Présentation des données

Nous rappelons la carte du magnétisme lorsqu'il n'est pas filtré des perturbations .

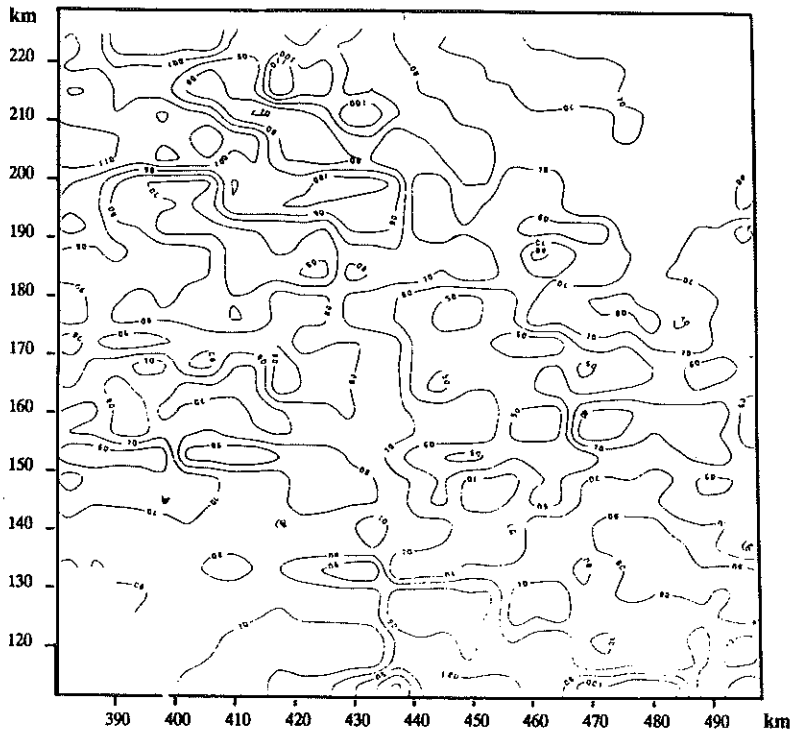


Figure 4.1 : Magnétisme spatio-temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes

4 .2 .2 Les modèles disponibles

Les modèles disponibles sont au nombre de 3 . Chaque estimateur D*(.) sera étiqueté par le n° du modèle qui lui est associé . On rappelle ces modèles :

- $\{ \mathcal{V}_2^x ; k = 2 ; T = 24^h ; K^1(h) \}$ avec $K^1(h) = a_1 |h|$ $\longrightarrow$ $D_1^*(t)$
- $\{ \mathcal{V}_3^x ; k = 2 ; T = 24^h ; K^2(h) \}$ avec $K^2(h) = a_\beta$ pépite$(h) + a_3 |h|^3$ $\longrightarrow$ $D_2^*(t)$
- $\{ \mathcal{V}_4^{x\,t} ; T = 24^h, 12^h ; k = 2 ; K^3(h) \}_{prior}$ avec $K^3(h) = a_2 |h|^2 \text{ Log}(|h|)$ $\longrightarrow$ $D_3^*(t)$

On rappelle que l'indice "prior" signifie que le paramètre (ou l'ensemble) auquel il est affecté est arbitraire .

4 .2 .3 Estimation avec le premier modèle

Le modèle est :

$\{ \mathcal{V}_2^x \ ; \ k = 2 \ ; \ T = 24^h \ ; \ K^1(h) \}$ avec $K^1(h) = a_1 |h|$ et $a_1 = -1.02$

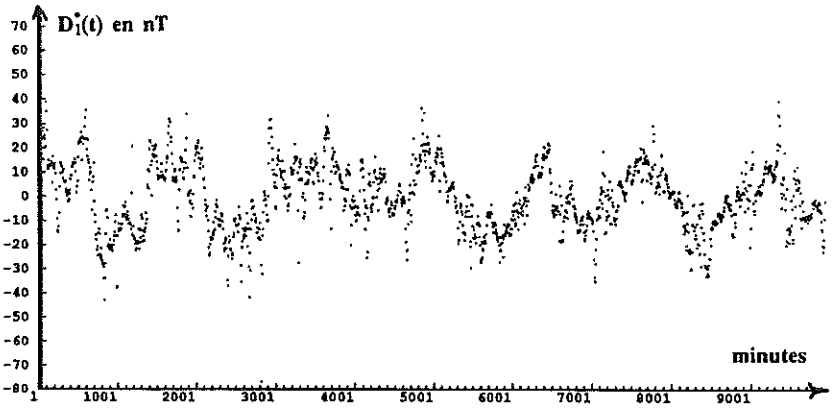


Figure 4.2 : Perturbations estimées (premier modèle)

L'estimateur obtenu (figure 4.2) présente les particularités suivantes :

- Il manifeste une vague tendance à l'auto-répétition .
- Il est fortement bruité .

On explique son bruit par l'incompatibilité entre la géométrie $\mathcal{V}_2^x$ (approximativement une droite de points) et le degré k = 2 de la FAI spatiale .

La figure 4.3 (courbe du haut) représente le spectre d'amplitude associé à cet estimateur . Son examen confirme l'analyse précédente, à savoir :

- La présence d'une forte amplitude à 24 ^{h} (10,2 nT^2) .
- Un spectre fortement bruité .

Outre la raison invoquée précédemment, il semblerait également que ces hautes fréquences de l'estimateur soient liées aux brusques variations locales de y(x). En effet :

- Krakatao se caractérise par une forte variabilité de Y(x) ,
- la géométrie $\mathcal{V}_2^x$ se caractérise par son équivalence avec une droite de points équidistants de 1 ^{h} de navigation (la distance inter-profil) et orthogonale aux profils,.

Or les profils sont ici parallèles à la direction principale de la faille , ce qui rend la droite des points de $\mathcal{V}_2^x$ orthogonale à celle-ci . Il en résulte que $\mathcal{V}_2^x$ (à x fixé) échantillonne des valeurs peu homogènes, très fluctuantes et avec un pas d'échantillonnage inadapté au spectre de Y(x) , ce qui peut expliquer ces fluctuations .

Afin de mieux distinguer les basses fréquences de $D_1^*(.)$, une coupure spectrale de ses énergies a été faite à la période de coupure $T_c = 125^{mn}$.
La figure 4.3 (courbe du bas) montre les évolutions des basses fréquences de l'estimateur. L'estimateur coupé n'est montré qu'à titre indicatif et ne servira pas au filtrage .

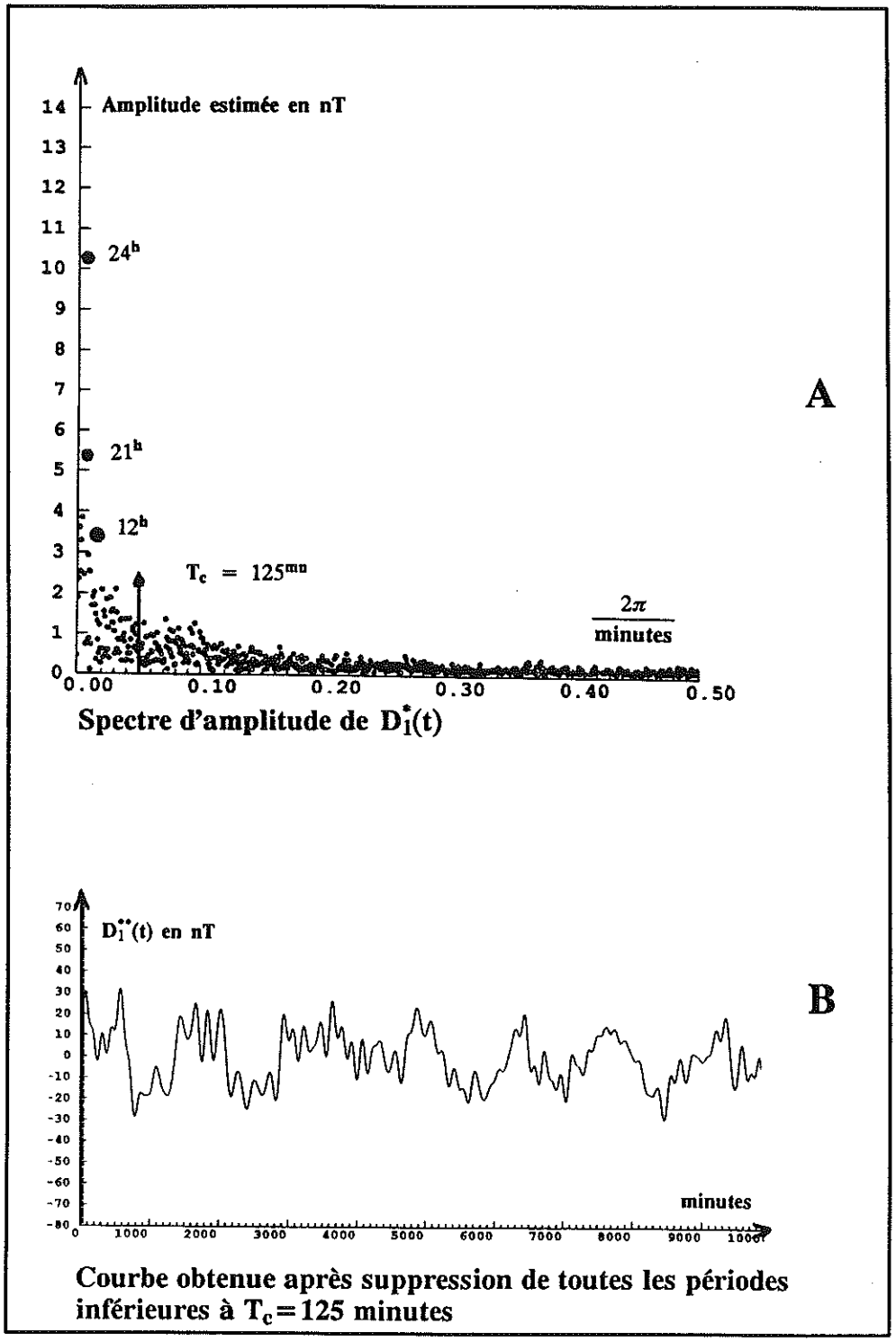


Figure 4.3 : Spectre de l'estimateur et post-filtrage (premier modèle)

La figure 4.4 représente la carte spatiale des anomalies estimées $Y_1^*(x)$ après l'opération de filtrage :

$$Y^*(x) = z(x,t) - D^*(t) \qquad (4.1)$$

appliquée avec $D^*(.) = D_1(.)$

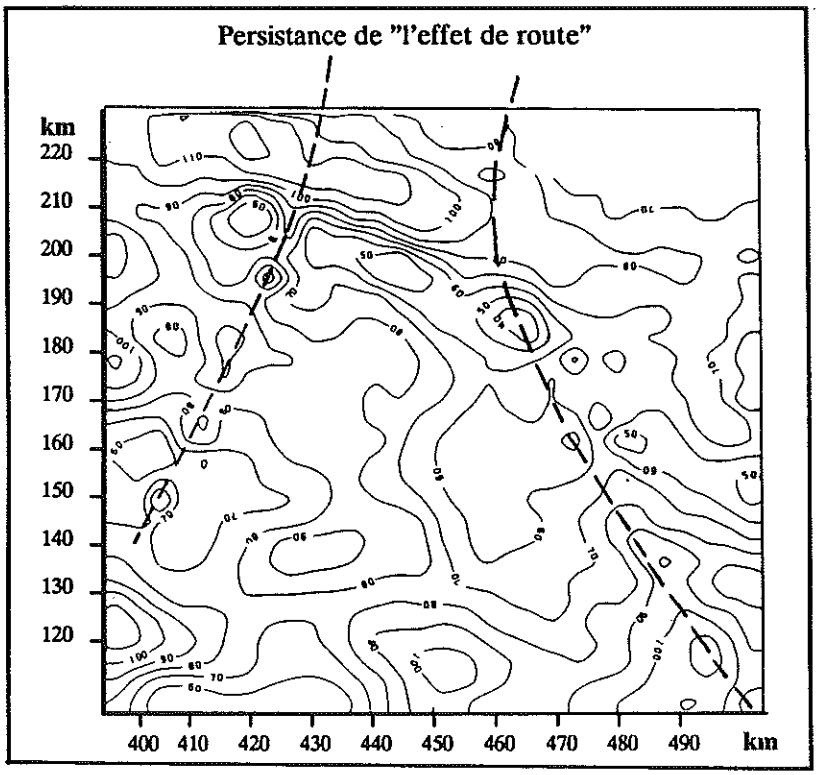


Figure 4.4 : Anomalies spatiales estimées (premier modèle)

Lorsqu'elle est comparée à celle du magnétisme avant filtrage (figure 4.1) cette carte montre le gain de précision apporté . On note en effet l'apparition de structures spatiales qui correspondent à la géologie du lieu . Les commentaires du géophysicien (cf conclusions 4.2.6) confirmeront cette remarque . On remarquera de plus qu'il subsiste un "effet de route" dû aux profils obliques qui contiennent une information incohérente par rapport à celle des profils qu'ils recoupent. Ceci signifie qu'en cet endroit, le filtrage est insuffisant .

Conclusion sur le premier modèle :

Il a engendré un estimateur temporel fortement bruité . Malgré cela, l'estimateur $Y_1^*(x)$ obtenu constitue une nette amélioration par rapport à une cartographie sans filtrage . Si le bruit de $D_1^*(t)$ n'apparait pas dans $Y_1^*(x)$, c'est essentiellement parce que la cartographie spatiale est précédée d'une estimation de $Y_1^*(x)$ aux noeuds d'une maille régulière de pas sensiblement égal à la distance inter-profil. Comme cette opération de maillage équivaut à un filtrage des hautes fréquences que $Y_1^*(x)$ aurait pu embarquer dans l'opération 4.1, l'estimateur résultant $Y_1^*(x)$ est beaucoup plus régulier (dans l'espace) que ne l'est $D_1^*(t)$ (dans le temps) .

4 .2 .4 Estimation avec le deuxième modèle

Le modèle utilisé est :

$\{ \mathcal{V}_3^x ; k = 2 ; T = 24^h ; K^2(h) \}$ avec $K^2(h) = a_\beta$ pépite(h) + $a_3 |h|^3$, $a_\beta = 2.0190$ et $a_3 = 0.1672$

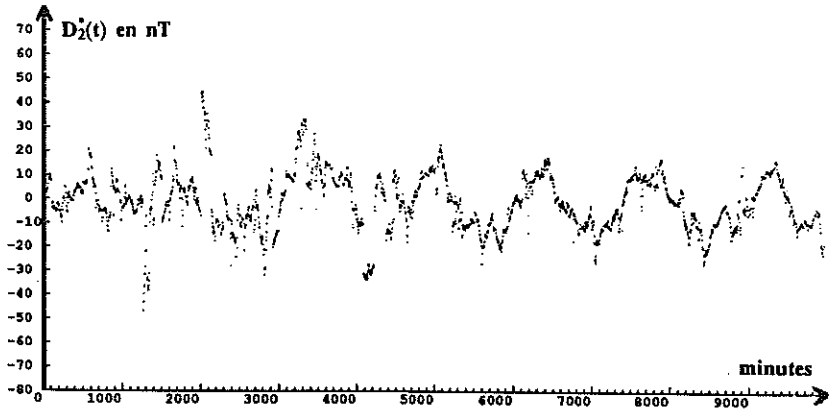


Figure 4.5 : Perturbations estimées (deuxième modèle)

L'estimateur obtenu (figure 4.5) présente les particularités suivantes :

- Il est moins bruité que précédemment .
- Sa tendance périodique est plus marquée .
- Son domaine de variation est plus important que précédemment .

La figure 4.6 (courbe du haut) représente le spectre d'amplitude associé à cet estimateur . Son examen confirme l'analyse précédente, à savoir :

- La présence d'une amplitude plus importante à 24 ^{h} (11,5 nT^2) .
- Un spectre dont le bruit est nettement atténué .

Cette amélioration est due à :

- La géométrie V_3^x qui est adaptée à k=2 .
- Un diamètre de voisinage plus important, puisqu'il est ici en moyenne de 30 ^{km} alors qu'il était de 9 ^{km} pour $\mathcal{V}_2^x$.
  Il en résulte que nous nous trouvons ici plutôt dans le cas d'une estimation à voisinage unique .
  De ce fait, les fréquences embarquées par la pratique glissante se déplacent vers l'infra-rouge et l'estimateur résultant est moins bruité .

Afin de mieux distinguer les basses fréquences de $D_2^*(.)$, nous avons effectué une coupure spectrale de ses énergies à la période de coupure $T_c = 125$ ^{mn} . La figure 4.6 (courbe du bas) montre les évolutions des basses fréquences de l'estimateur .

L'estimateur coupé n'est montré qu'à titre indicatif et ne servira pas au filtrage .

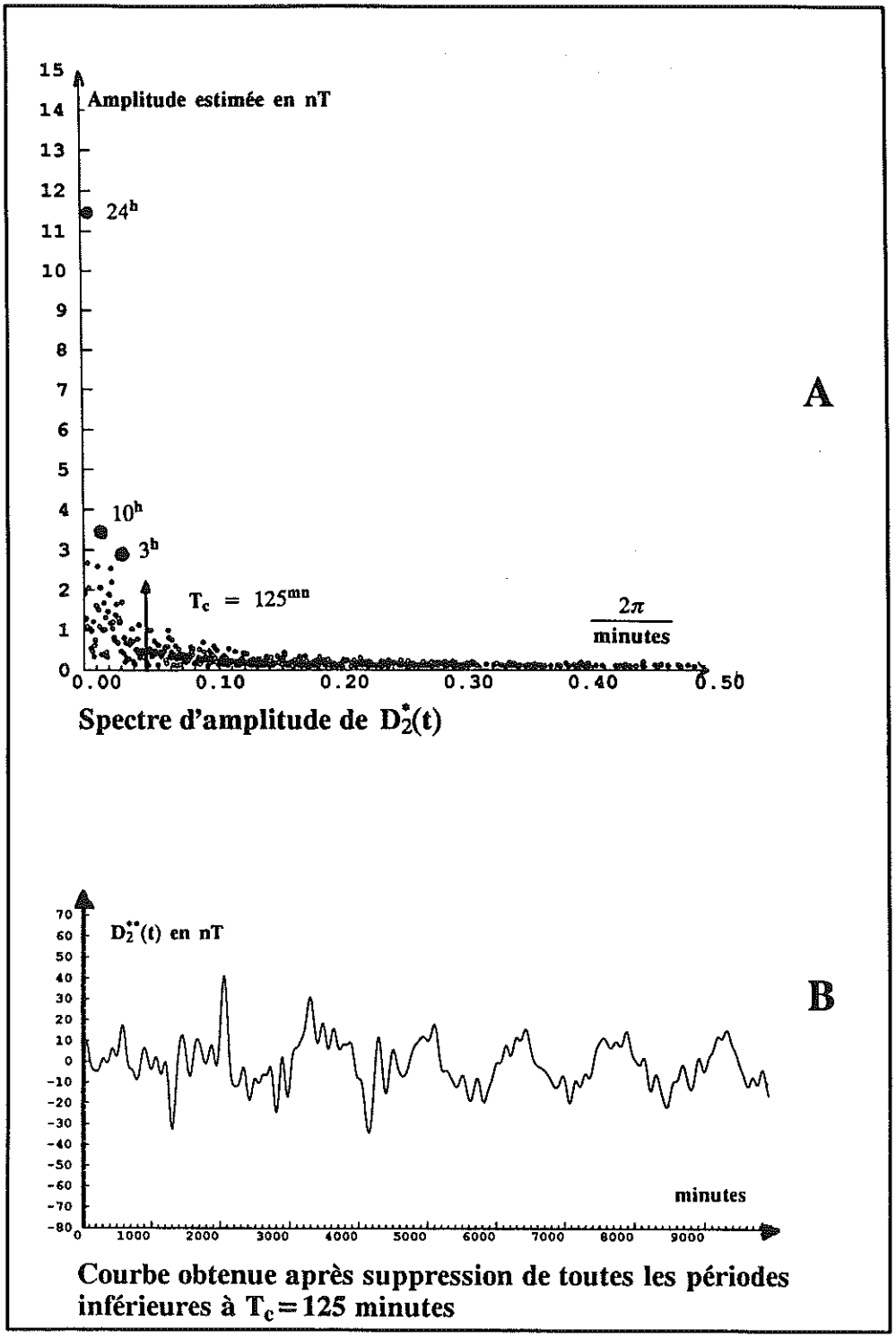


Figure 4.6 : Spectre de l'estimateur et post-filtrage (deuxième modèle)

La figure 4.7 représente la carte spatiale des anomalies estimées $Y_2^*(x)$ après l'opération de filtrage 4.1 .

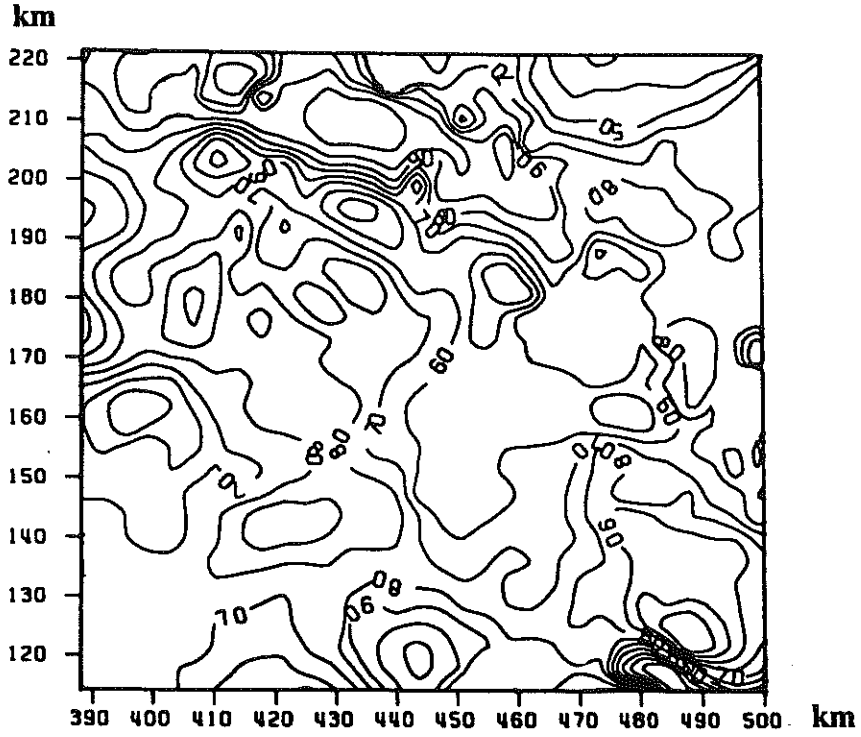


Figure 4.7 : Anomalies spatiales estimées (deuxième modèle)

Comparativement à la carte de $Y_1^*(x)$, on note :

- Une similarité pour les tendances à grande échelle .
- Des isolignes de $Y_2^*(x)$ plus torturées que celles de $Y_1^*(x)$.
- Un renforcement de "l'effet de route" .

Conclusion sur le deuxième modèle :

Il semblerait que la régularité obtenue dans le temps se fasse au détriment de celle de l'estimateur spatial résultant. Cette régularité a été obtenue en augmentant le diamètre de $\mathcal{V}$, et en faisant tendre notre pratique vers une estimation à voisinage unique . Par cette opération, on a obtenu un estimateur qui ne possède plus assez de détail.

A priori, c'est entre $Y_1^*(x)$ et $Y_2^*(x)$ que la réalité semble se situer . Mais cette dernière remarque est quelque peu arbitraire, et on consultera l'analyse du Géophysicien (cf conclusion) pour se faire une opinion .

4 .2 .5 Estimation avec le troisième modèle

Le modèle utilisé est :

$$\left[\mathcal{V}_4^{x\,t} \; ; \; T = 24^h, 12^h \; ; \; k = 2 \; ; \; K^3(h) \right]_{prior} \text{ avec } K^3(h) = a_2 \, |h|^2 \, \text{Log}(|h|) \text{ et } a_2 = 1.25$$

Ce modèle possède 3 particularités :

- Il est intégralement "a-priori" .
- Il est affecté à une géométrie spatio-temporelle $\mathcal{V}_4^{x\ t}$
- Il contient également la période 12 ^{h} .

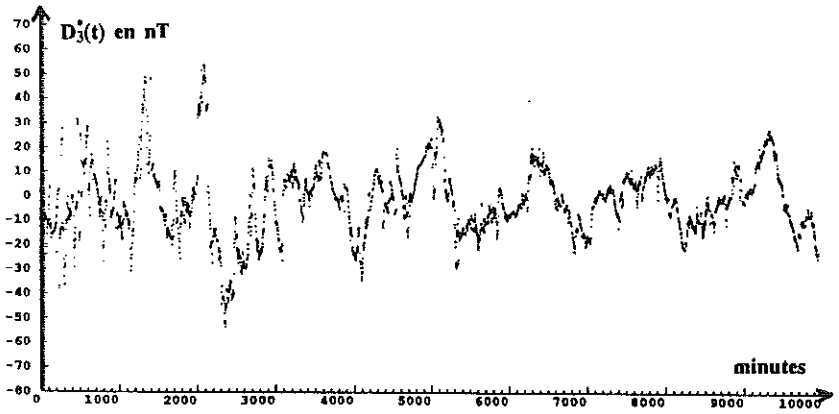


Figure 4.8 : Perturbations estimées (troisième modèle)

L'estimateur obtenu (figure 4.8) présente les particularités suivantes :

- Son domaine de variation est plus important que précédemment .
- Si les 3 dernières périodes sont comparables à celles de $D_1^*(.)$ et $D_2^*(.)$, il n'en est pas de même pour les 2 premiers jours où $D_3^*(.)$ présente de nombreuses "explosions" caractéristiques de l'instabilité numérique du système .

La figure 4.9 (courbe du haut) représente le spectre d'amplitude associé à cet estimateur . On remarque que :

- L'amplitude à 24 ^{h} s'est réduite à 9.5 nT^2 (11,5 nT^2 dans le cas précédent).
- Il y a émergence d'une amplitude à 12 ^{h} (8 nT^2)

Le gain énergétique de $D_3^*(.)$ est donc essentiellement dû à l'énergie captée par la période 12^h.

Afin de mieux distinguer les basses fréquences de $D_3^*(.)$, nous avons effectué une coupure spectrale de ses énergies à la période de coupure Tc = 125 ^{mn}. La figure 4.9 (courbe du bas) montre les évolutions des basses fréquences de l'estimateur .
L'estimateur coupé n'est montré qu'à titre indicatif et ne servira pas au filtrage .

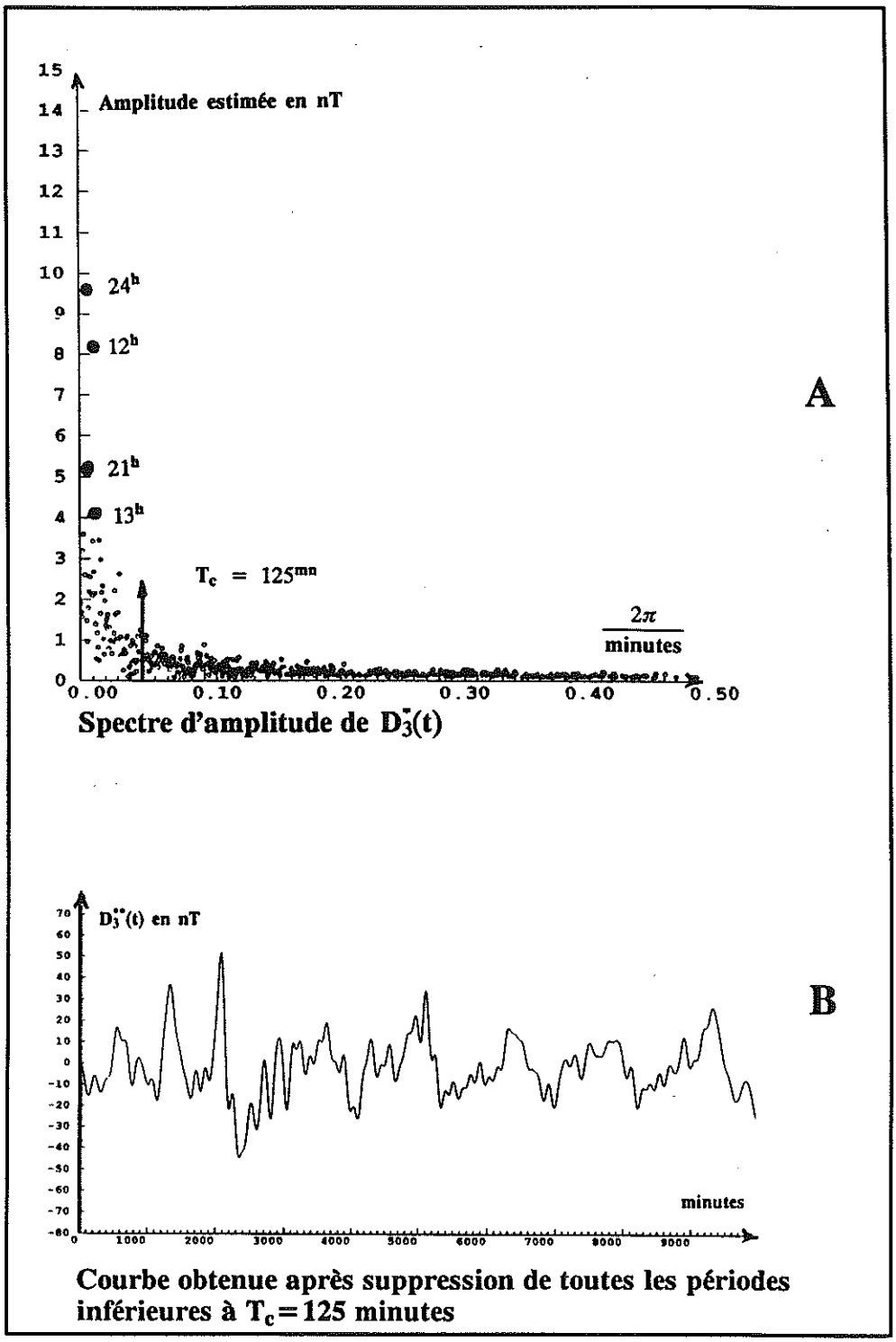


Figure 4.9 : Spectre de l'estimateur et post-filtrage (troisième modèle)

La figure 4.10 représente la carte spatiale des anomalies estimées $Y_3^*(x)$ après l'opération de filtrage 4.1 .

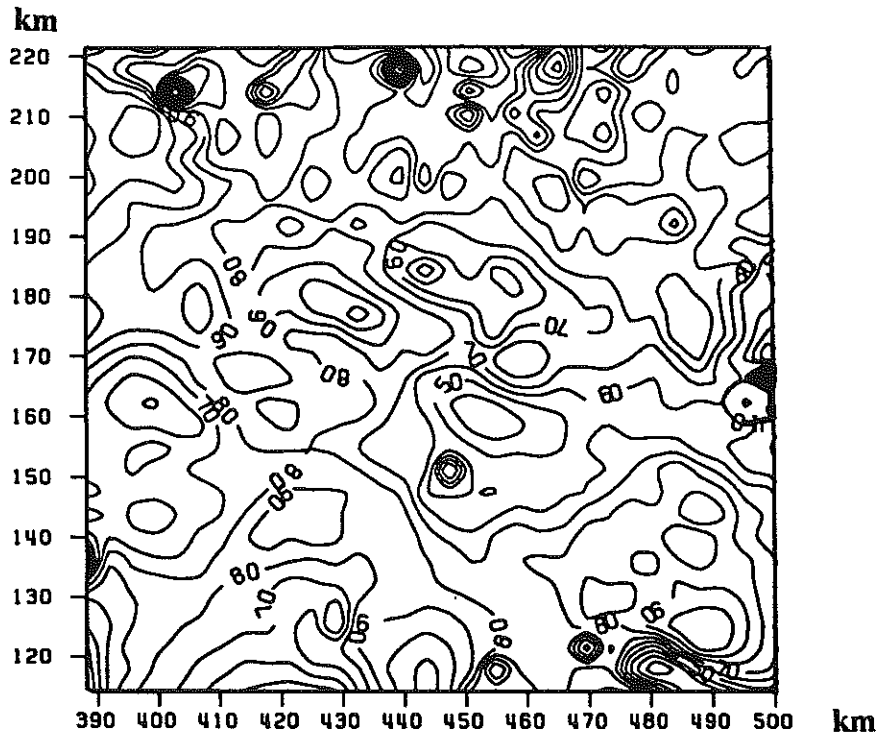


Figure 4.10 : Anomalies spatiales estimées (troisième modèle)

Comparativement aux cartes précédentes on note :

- La présence de plusieurs zones instables .
  Ce sont celles associées aux explosions de $D_3^*(.)$.
- De nombreuses différences, et seuls quelques traits structuraux subsistent .

Conclusion sur le troisième modèle :

Il ne figurait ici qu'à titre indicatif, puisque son inférence n'avait pas été possible. La raison essentielle de son inadéquation aux données est que la géométrie $\mathcal{V}_4^{x\,t}$ rend instable le système de Krigeage (cf chapitre 3) .

Outre les instabilités chroniques qu'elle engendre, cette géométrie n'a pas pu rendre les services escomptés, car le diamètre temporel de 10 ^{h} qu'elle contient rend impossible la séparation entre ce qui dans D(t) appartient à la période 24 ^{h} de ce qui appartient à la période 12 ^{h} .

Si on identifie le KT à un balayage spectral constitué de ces deux périodes 24^h et 12^h, il s'est alors passé la chose suivante :

- Lorsque le balayage était focalisé sur la période 24 ^{h} et qu'un segment seulement de sinusoïde de longueur 10 ^{h} analysait le signal , le système captait quelque chose de moyen entre ce qui est réellement dû à cette période, et une partie de l'énergie associée à 12 ^{h} .
- Il en était de même lorsque le balayage était focalisé sur la période 12 ^{h} .

En conséquence, c'est la même énergie qui est estimée deux fois, d'où cet estimateur .

4 .2 .6 Conclusions sur Krakatao

Bilan :

La figure 4.11 montre le chemin parcouru par le filtrage . L'absence de vérifications directes ne facilite pas l'élaboration de conclusions . Il est cependant possible de résumer un certain nombre de remarques :

- Le premier modèle engendre une courbe bruitée dans le temps, mais un estimateur régulier dans l'espace .
- Le deuxième modèle engendre une courbe artificiellement régulière dans le temps, puisqu'engendrant des irrégularités dans l'espace .
- La configuration $\Psi_4^{x\,t}$ doit être définitivement écartée .

La parole est à un Géophysicien :

Nous remercions Ph. Huchon (ref [I–2]) de s'être prêté au jeu suivant : nous lui avons fourni la carte du magnétisme brut, celles des 3 estimations spatiales , ainsi que les courbes des séries estimées qui leur sont associées, et nous lui avons demandé de choisir, sans autres précisions de notre part .

Ci suit l'intégralité de son propos :

< < Les critères utilisables pour choisir entre les trois cartes proposées sont de deux ordres :

- objectifs ,
- subjectifs .

Les critères objectifs sont :

1) La coïncidence entre l'estimateur $D^*(t)$ de la variation diurne et la variation réelle D(t) enregistrée dans un observatoire (proche du lieu de la campagne) .

2) Les erreurs résiduelles aux points de croisement .

Ni l'un ni l'autre n'étant à ma disposition, je me garderai donc de conclure en prétendant que le choix fait est, au moins en partie, objectif...

Cependant, on peut noter que la carte $Y_2^*(x)$ et surtout la carte $Y_1^*(x)$ présentent d'importants effets de route sur les deux routes transverses à la boîte, ce qui traduit des erreurs aux points de croisements, erreurs que l'on souhaite a priori minimiser .

Les critères subjectifs dépendent quant à eux de la connaissance géologique que l'on a de la région considérée .

Si l'on considère l'aspect général de la carte, il est clair (en tout cas pour moi!) que les cartes $Y_1^*(x)$ et $Y_2^*(x)$ sont satisfaisantes, dans la mesure où l'on y voit des structures, c'est à dire des linéations magnétiques, en particulier la zone (en creux) d'orientation WNW–ESE dans la partie supérieure de la carte .

Cette zone a l'intérêt d'être parallèle à la structure de la marge continentale, et a donc, à mon avis, de bonnes chances d'être réelle, et non un artefact .

Il est plus difficile d'utiliser ce type de critère "géologique" dans les autres parties de la carte, sauf peut–être en ce qui concerne une amorce de linéations nord–sud dans la partie inférieure de la carte, qui est située en partie sur la fosse de Java et sur la plaque océanique indienne .

Un autre critère subjectif est lié à l'aspect de l'estimateur $D^*(t)$ de la variation diurne .

L'estimateur $D_3^*(t)$ est le plus erratique, avec des variations très brusques de l'amplitude, ce qui confirmerait la qualité plutôt médiocre attribuée à la carte $Y_3^*(x)$.

Choisir entre $D_1^*(t)$ et $D_2^*(t)$ est plus délicat : dans les deux cas, la périodicité est plutôt bonne; tout au plus, on note que $D_1^*(t)$ est plus bruité dans les hautes fréquences . (moins bonne stabilité de l'estimation ? car voisinage temporel plus petit ?)

En résumé, mon choix se porterait, concernant la carte à interpréter, sur $Y_1^*(x)$ ou $Y_2^*(x)$, avec peut–être une légère préférence pour $Y_2^*(x)$ > > .

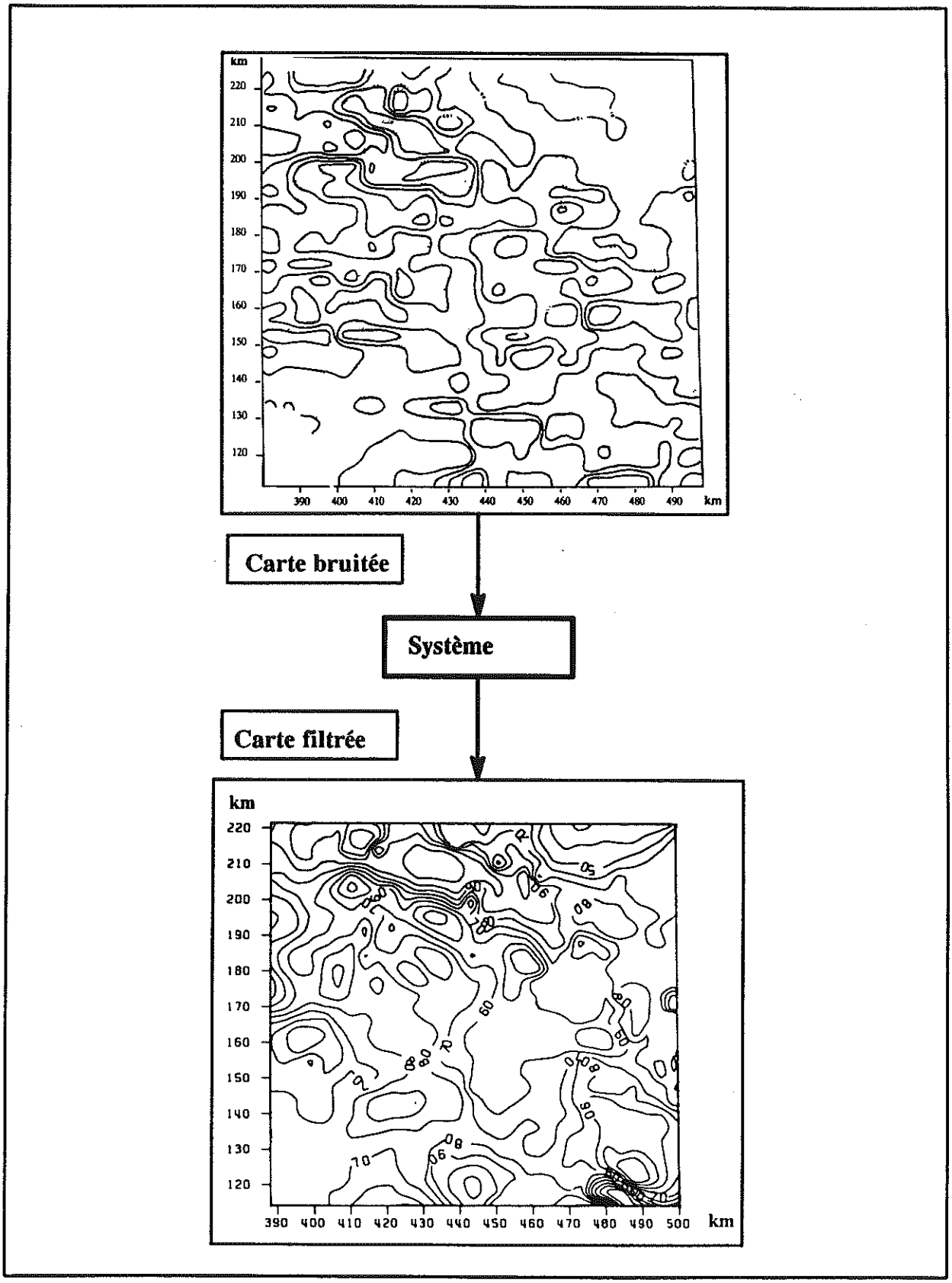


Figure 4.11 : Le chemin parcouru grâce au filtrage

4 .3 Filtrage des données de la campagne SEAPERC

Ce jeu de données est caractérisé par des périodes temporelles marquées et la possibilité de vérifier les filtrages puisque nous disposons de la série temporelle Huancayo qui représente une approximation acceptable de la réalité des perturbations sur le lieu de la campagne .

4 .3 .1 Présentation des données

Nous rappelons succinctement la carte du magnétisme lorsqu'il n'est pas filtré des perturbations (figure 4.12), l'allure des perturbations près de la zone d'étude (figure 4.13) et l'allure de la carte lorsque z(x,t) est corrigé par cette série temporelle Huancayo (figure 4.14).

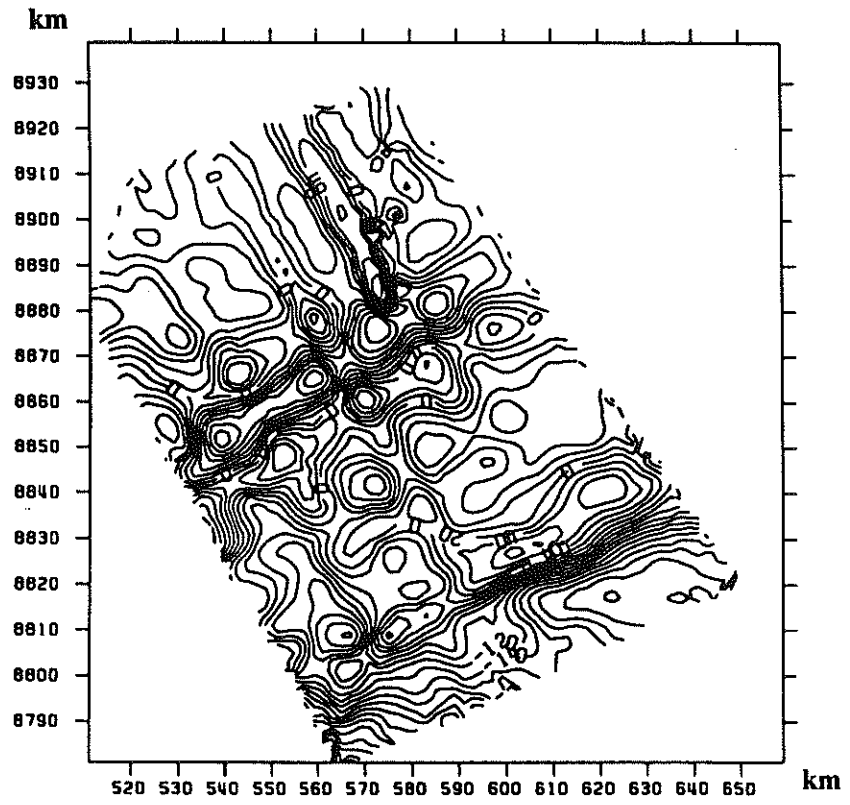


Figure 4.12 : Magnétisme spatio-temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes

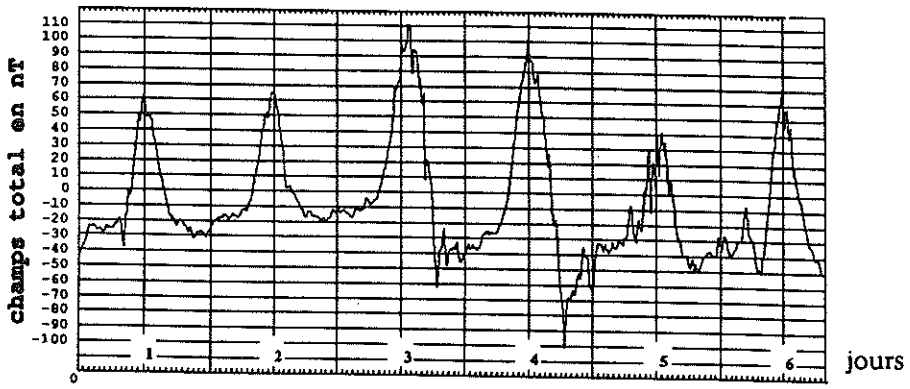


Figure 4.13 : Perturbations diurnes D(t) enregistrées à l'observatoire fixe de Huancayo (Pérou)

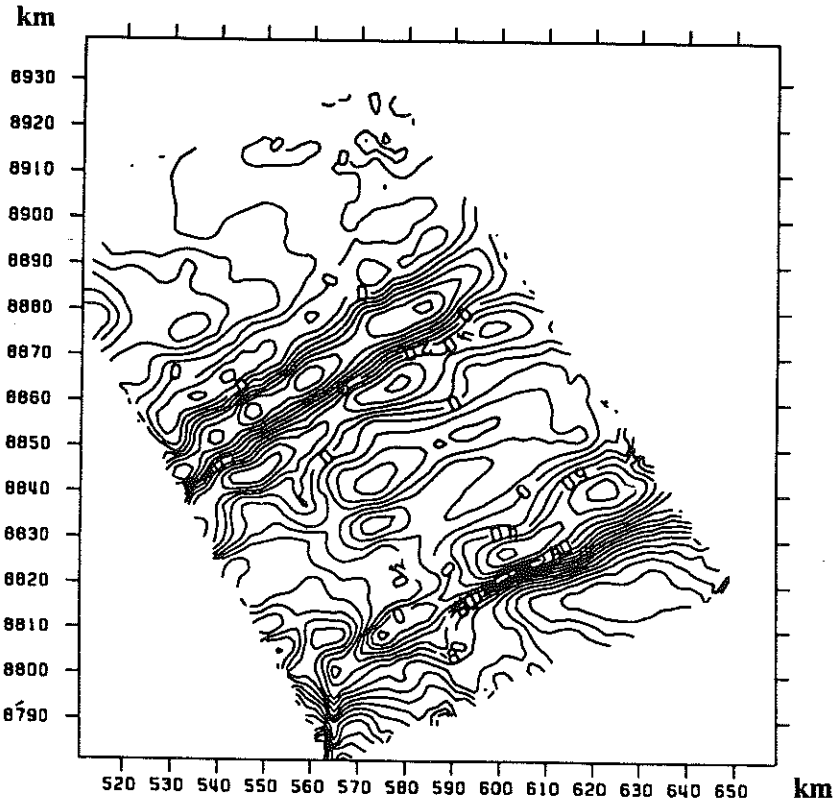


Figure 4.14 : Cartographie des anomalies spatiales y(x) = z(x,t) – D(t) où D(t) est la fonction perturbations diurnes à Huancayo

4 .3 .2 Les modèles disponibles

Les modèles disponibles sont aux nombre de 5 :

- $\{ \mathcal{V}_2^x ; k = 1 ; T_{prior} = 24^h ; K^1(h) \}$ avec $K^1(h) = a_2 \, h^2 \, Log(|h|)$ $\longrightarrow D_1^*(t)$

- $\{ \mathcal{V}_2^x ; k = 1 ; T_{prior} = 24^h, 12^h ; K^2(h) \}$ avec $K^2(h) = a_2 \, h^2 \, Log(|h|)$ $\longrightarrow D_2^*(t)$

- $\{ \mathcal{V}_3^x ; k = 1 ; T = 24^h, 12^h ; K^3(h) \}$ avec $K^3(h) = a_\beta$ pépite$(h) + a_2 \, h^2 \, Log(|h|)$ $\longrightarrow D_3^*(t)$

- $\left[\mathcal{V}_4^{x\,t} ; k_{prior} = 1 ; T_{prior} = 24^h, 12^h ; K^{prior} \right]$ avec $K^4(h) = K^{prior}(h) = K^3(h)$ $\longrightarrow D_4^*(t)$

- $\left[\mathcal{V}_4^{x\,t} ; k_{prior} = \emptyset ; T_{prior} = 24^h, 12^h ; K^5(h) \right]_{prior}$ avec $K^5(h) = a_1 \, |h|$ $\longrightarrow D_5^*(t)$

4 .3 .3 Estimation avec le premier modèle

Le modèle utilisé est :

$\{ \mathcal{V}_2^x ; k = 1 ; T_{prior} = 24^h ; K^1(h) \}$ avec $K^1(h) = a_2 \, h^2 \, Log(|h|)$ et $a_2 = 10.62$

Dans ce modèle, l'introduction d'une fréquence est arbitraire puisque celle-ci n'a pas pu être inférée lors de l'analyse et la modélisation .

La partie haute de la figure 4.15 contient les évolutions conjointes de $D_1^*(t)$ (trait pointillé) et de D(t) (trait continu) sur le champ temporel des mesures . L'absence d'estimation pour $t \in [3300 , 5800]$ correspond à une absence de mesure du magnétisme brut, et donc d'estimation . (Nous nous sommes refusé à utiliser les modèles en extrapolation) . L'estimateur $D_1^*(t)$ possède les spécificités suivantes :

- Il manifeste certaines explosions (les points notés A, B et C sur la figure) .
- Seules les 2 premières périodes de $D_1^*(t)$ suivent les évolutions de D(t) .

La partie basse de la figure constitue un agrandissement des deux premières périodes de $D_1^*(t)$ et D(t) .On note que $D_1^*(t)$ possède une ondulation parasitaire de période sensiblement égale à 1 ^{h} dont on attribue les causes à :

- L'effet "sautant " de la géométrie V_2^x
- Un recouvrement spectral (ref **[TdS-5]**) lié à un pas d'échantillonnage (1 ^{h}) inadapté au phénomène .

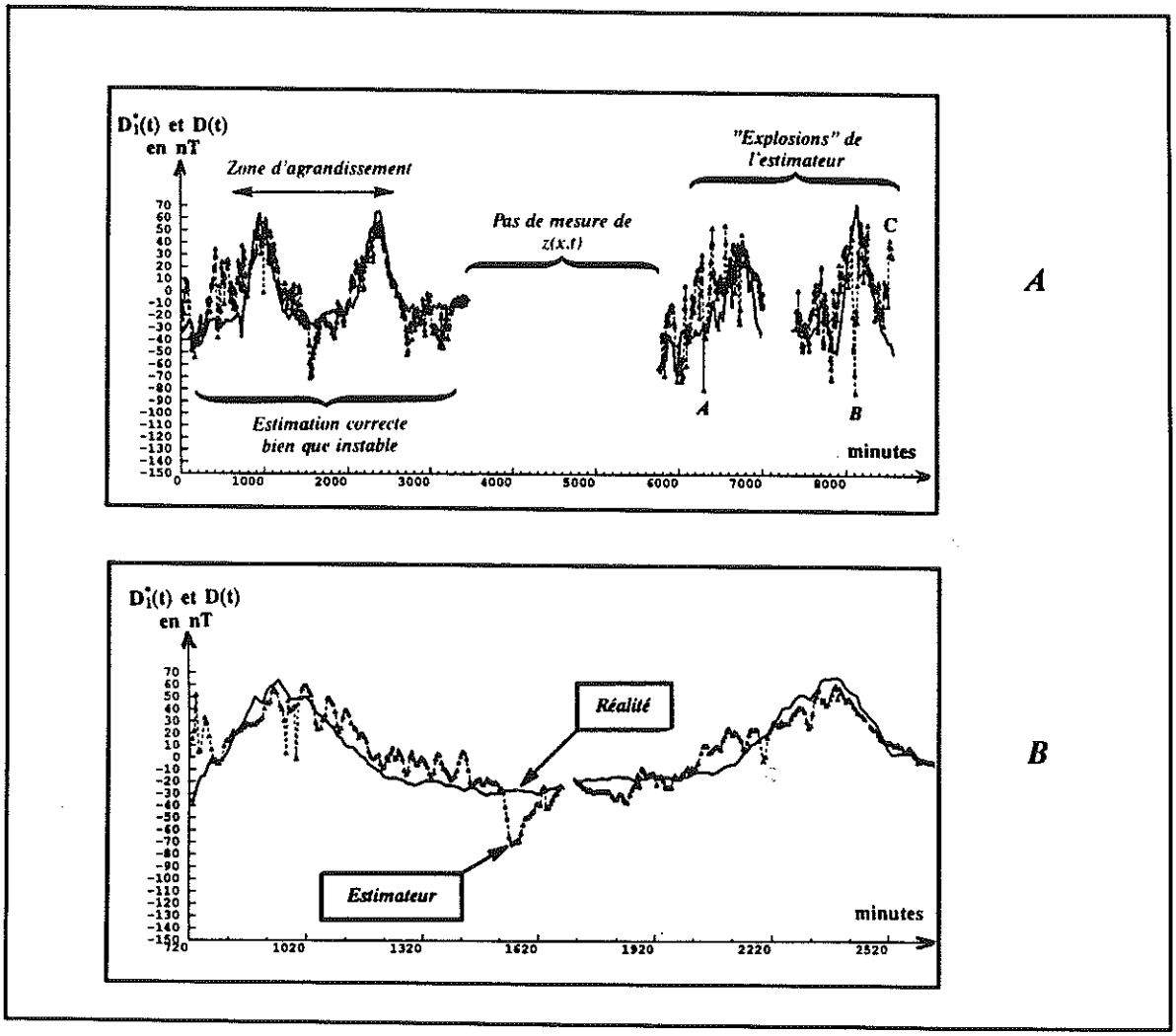


Figure 4.15 : Estimateur et réalité des perturbations (premier modèle)

Les statistiques élaborées à partir de l'estimateur et de la réalité donnent :

- Nombre de points estimés = 5395
- Erreur moyenne $E[\ D_1^*(.) - D(.)\ ] = -6.78$ nT
- Variance expérimentale $Var[\ D_1^*(.) - D(.)\ ] = 577.93\ nT^2$

A propos de ces statistiques, il est opportun ici de mentionner des expériences mono-dimensionnelles effectuées sur la série temporelle Huancayo et qui ne figurent pas dans cette thèse. Elles consistaient à introduire dans un système de KT des extraits de série temporelle dans le but de compléter l'information par un estimateur. La covariance utilisée était un modèle pépitique. Comme nous disposions de l'information complète, nous avions pu comparer estimation et réalité et de ce fait établir des variances expérimentales sur le même champ temporel qu'ici. Lorsque le diamètre valait 30 h, la variance expérimentale obtenue était de 560 nT^2. Ceci est remarquablement proche de la variance de 578 nT^2 obtenue ici, d'autant plus que le diamètre temporel résultant de la géométrie γ_2^x est approximativement de 30h également.

Notre modèle spatio-temporel est donc pour ce diamètre, aussi efficace que si les données d'entrée étaient des échantillons de la courbe D(.).

La figure 4.16 représente la carte spatiale des anomalies estimées $Y_1^*(x)$ après l'opération de filtrage.

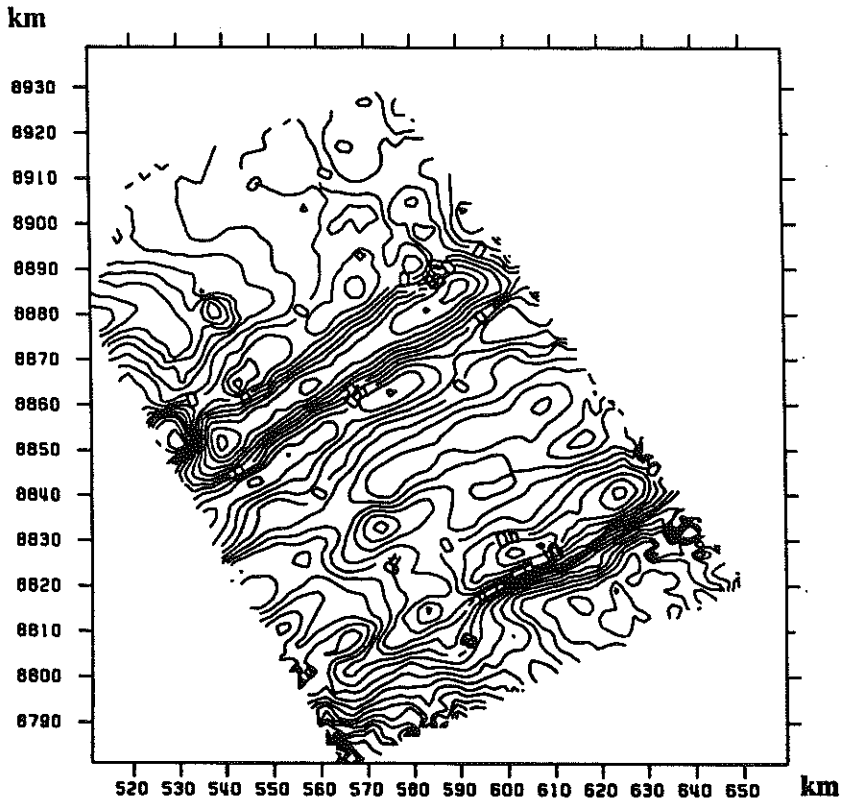


Figure 4.16 : Anomalies spatiales estimées (premier modèle)

Comparativement aux cartes des anomalies spatiales (figure 4.14) et du magnétisme non corrigé des perturbations (figure 4.12), on note :

- La disparition de "l'effet de route" .
- L'apparition de structures NO–SE .
- La présence de 3 zones à fort gradient qui n'existent pas dans la réalité . Ces zones sont celles associées aux explosions de $D_1^*(.)$.

Conclusion sur le premier modèle :

$D_1^*(.)$ est un estimateur acceptable, malgré une géométrie $\mathcal{V}_2^x$ peu élaborée et quelques explosions liées à des instabilités numériques du système .

A l'aide de ce modèle, fruste par son diamètre, son pas d'échantillonnage et son absence de covariance temporelle, on a su séparer convenablement l'énergie temporelle de l'énergie spatiale et obtenir une variance expérimentale de l'ordre de grandeur de celle obtenue lorsque le signal d'entrée est D(.) .

4 .3 .4 Estimation avec le deuxième modèle

Le modèle utilisé est :

$$\left\{ \mathcal{V}_2^x \ ; \ k = 1 \ ; \ T_{prior} = 24^h, 12^h \ ; \ K^2(h) \right\} \text{ avec } K^2(h) = a_2 \ h^2 \ \text{Log}(|h|) \text{ et } a_2 = 6.188$$

Dans ce modèle, l'introduction des fréquences est arbitraire.

De plus, on sait que la géométrie utilisée est incompatible avec la période 12^h . L'analyse des estimations résultantes va donc permettre de quantifier l'impact sur l'estimation d'une telle inadéquation . Les résultats seront d'autant plus démonstratifs que la CG utilisée est la même que dans le premier cas .

Les évolutions comparées de $D_2^*(.)$ et D(.) (figure 4.17) se passent de commentaires tant les explosions sont nombreuses et l'estimation erratique .

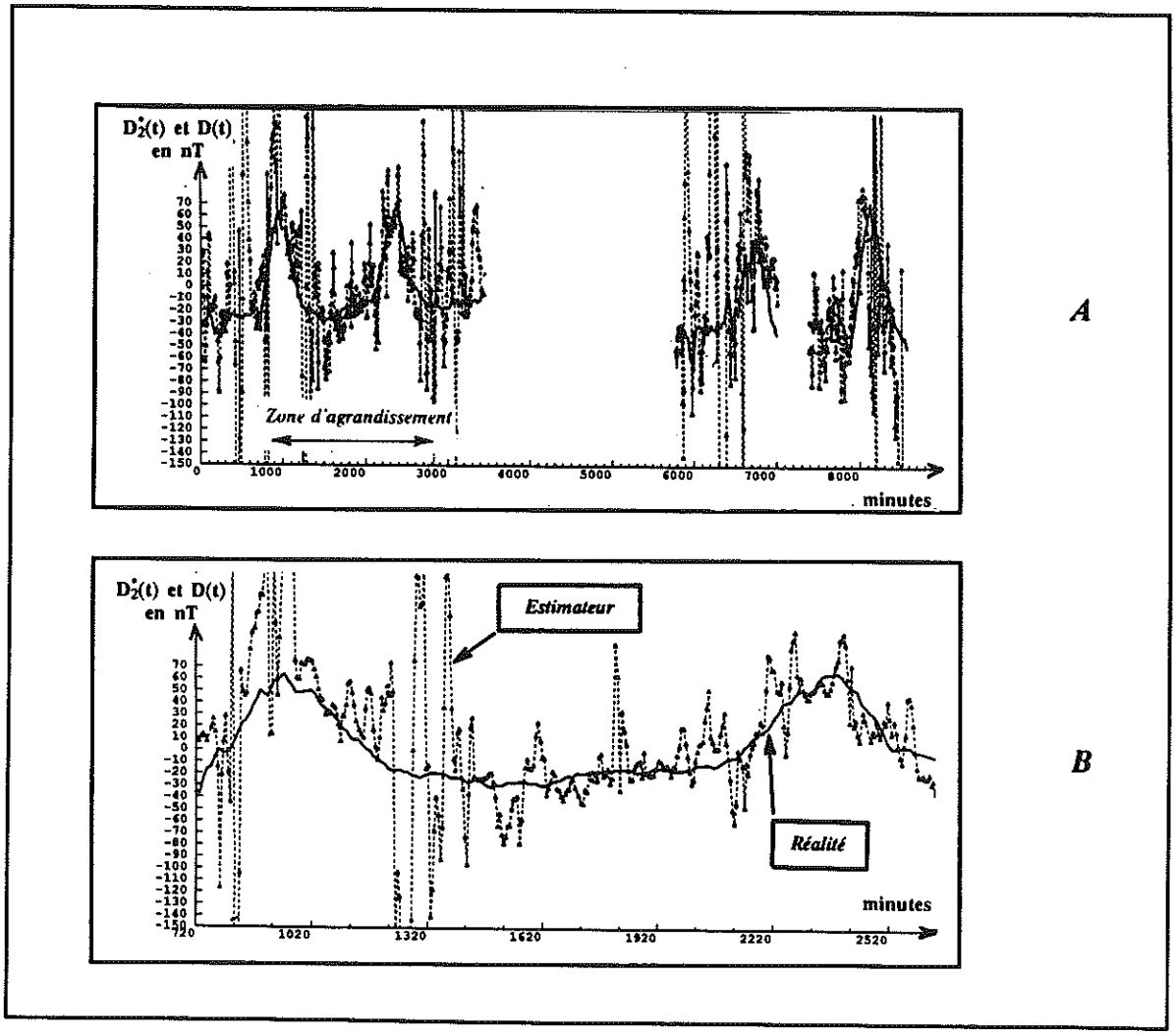


Figure 4.17 : Estimateur et réalité des perturbations (deuxième modèle)

Les statistiques résultantes reflètent bien ce comportement :

- Nombre de points estimés = 5395
- Erreur moyenne $E[\ D_2^*(.) - D(.)\ ] = -2.447$ nT
- Distance quadratique $E[\ D_2^*(.) - D(.)\ ]^2 = 18884$ nT^2
- Variance expérimentale $Var[\ D_2^*(.) - D(.)\ ] = 18878$ nT^2

Etant donné le caractère erratique de ce deuxième estimateur, nous n'avons pas jugé utile de présenter la carte spatiale résultante .

Conclusion sur le deuxième modèle :

La seule chose qui différencie l'élaboration de $D_2^*(.)$ de celle de $D_1^*(.)$, c'est l'introduction de la période 12^h dans le système . L'erraticité que cette modification entraîne, est liée à un pas d'échantillonnage (dans le temps et résultant de la géométrie spatiale) inadapté à cette période .

4 .3 .5 Estimation avec le troisième modèle

Le modèle utilisé est :

$$\left\{ \mathcal{V}_3^x \ ;\ k = 1 \ ;\ T = 24^h, 12^h \ ;\ K^3(h) \right\} \text{ avec } K^3(h) = a_\beta \text{ pépite}(h) + a_2\ h^2\ \text{Log}(|h|)\ ,$$
$$a_\beta = 0.213 \text{ et } a_2 = 6.16$$

Comme précédemment, il contient 2 périodes, mais il est associé à une géométrie de type $\mathcal{V}_3^x$ dont on rappelle qu'elle est équivalente à une maille spatiale de pas 1^h (de navigation) .On rappelle également que ce modèle a été intégralement inféré .

L'estimateur $D_3^*(t)$ (figure 4.18) possède les propriétés suivantes :

- Il ne manifeste pratiquement plus d'explosion .
- Toutes ses périodes sont comparables à celles de la réalité .

La partie basse de la figure constitue un agrandissement des deux premières périodes de $D_3^*(t)$ et D(t) . On note que $D_3^*(t)$ possède deux types d'oscillations parasitaires :

- Des parasites de périodes 1^h (le pas d'échantillonnage).
  Ils sont dus à un effet de recouvrement .
- Des parasites de périodes 400^{mn} .
  Ils sont dus à l'effet "sautant " très marqué de cette géométrie .
  En effet, cette dernière période correspond au diamètre moyen (exprimé en temps) de $\mathcal{V}_3^x$.

Les statistiques, élaborées à partir de l'estimateur et de la réalité donnent :

- Nombre de points estimés = 5395
- Erreur moyenne $E[\ D_3^*(.) - D(.)\ ] = -10.80$ nT
- Variance expérimentale $Var[\ D_3^*(.) - D(.)\ ] = 238.65$ nT^2

On constate une nette amélioration par rapport aux deux modèles précédents. Ces résultats sont une fois de plus comparables (pour le même diamètre) à ceux issus des expériences mono-dimensionnelles évoquées dans la section 4.3.3.

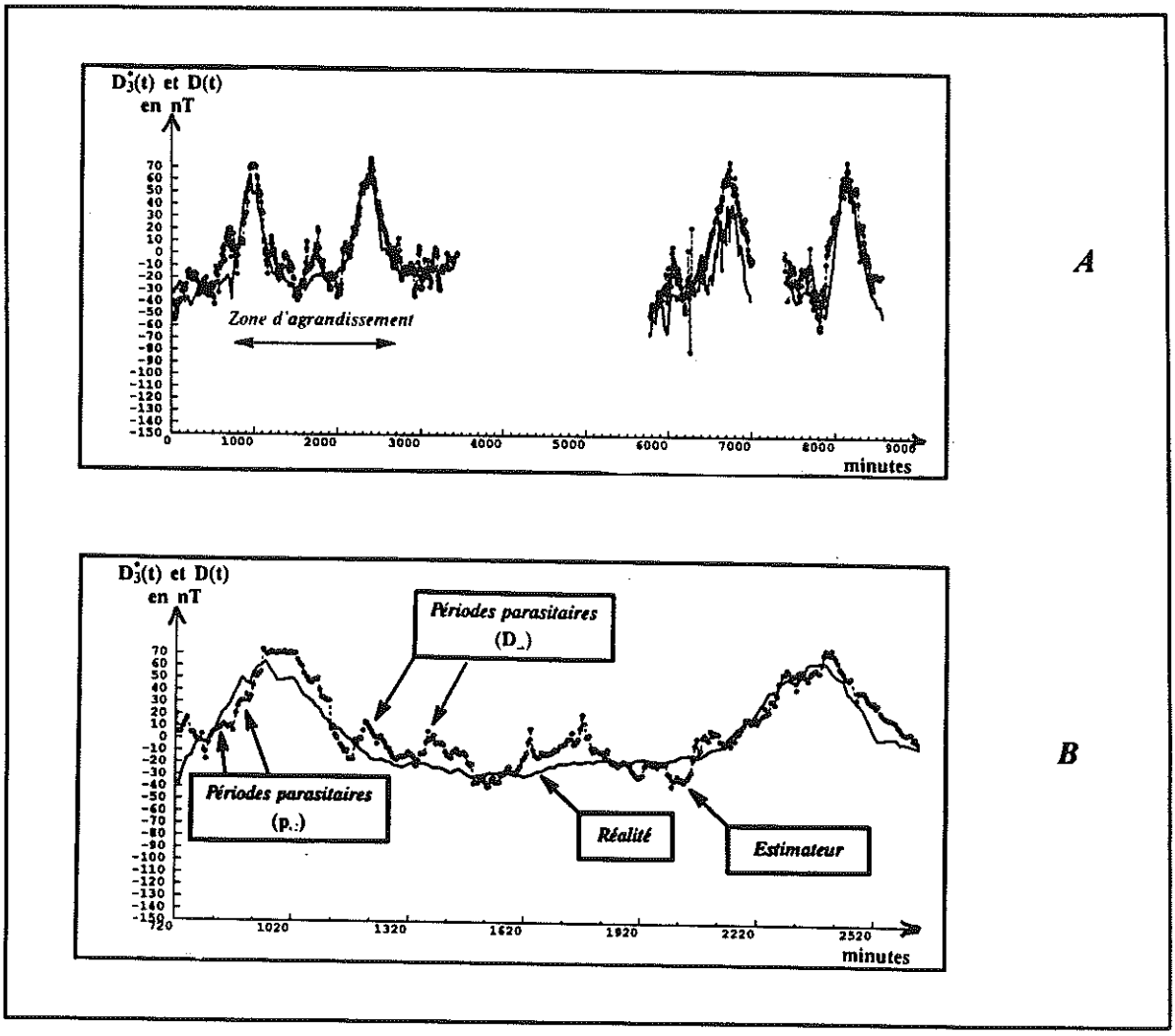


Figure 4.18 : Estimateur et réalité des perturbations (troisième modèle)

La figure 4.19 représente la carte spatiale des anomalies estimées $Y_3^*(x)$ après l'opération de filtrage 4.1 : Comparativement à la carte de $Y_1^*(x)$ (figure 4.16) et à la carte idéale, on note :

- La disparition des 3 zones erratiques .
- Une grande ressemblance avec la carte vraie .

Conclusion sur le troisième modèle :

Cet estimateur est très satisfaisant , ce qui a été rendu possible par :

- Le changement de géométrie qui a permis d'estimer correctement l'énergie associée à 12^h .
- L'introduction de la période 12^h qui a permis de mieux modéliser D(.) .
  On sait en effet de par l'étude de la série Huancayo (cf. chap. 2), que cette période est également très importante .

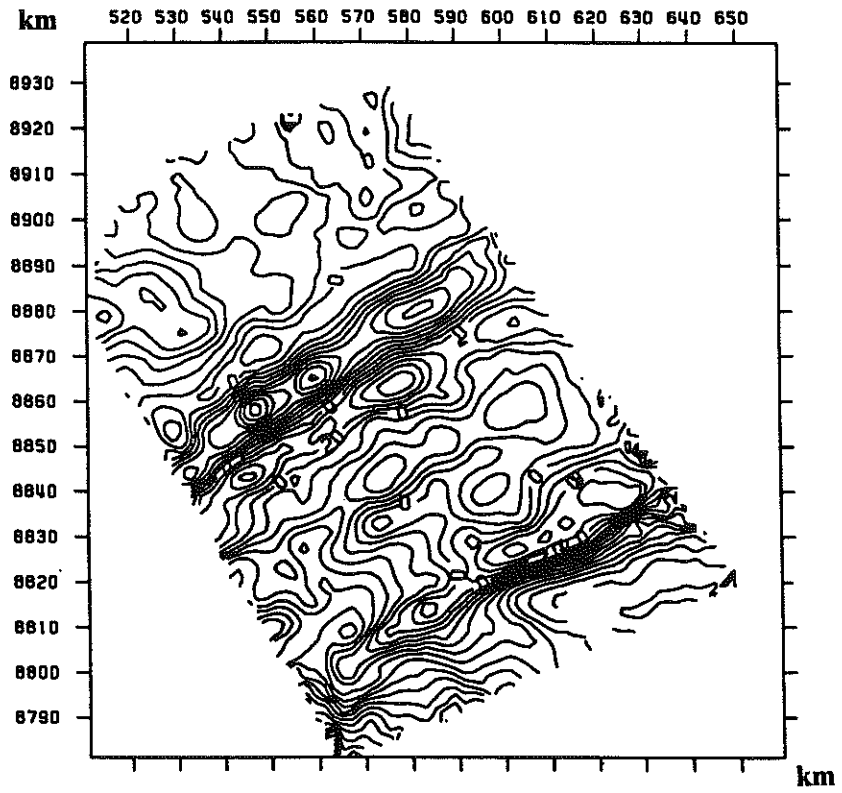


Figure 4.19 : Anomalies spatiales estimées (troisième modèle)

4 .3 .6 Estimation avec le quatrième modèle

Le modèle utilisé est :

$$\left\{ \mathcal{V}_4^{x\,t} \;;\; k_{prior} = 1 \;;\; T_{prior} = 24^h, 12^h \;;\; K^{prior}(h) = K^3(h) \right\}_{prior}$$

On rappelle que ce modèle est intégralement "a-priori"

La figure 4.20 représente un estimateur globalement comparable à $D_3^*(.)$, mais qui dans le détail se différencie par des explosions .

Les statistiques, élaborées à partir de l'estimateur et de la réalité, donnent :

- Nombre de points estimés = 5395
- Erreur moyenne $E[\, D_4^*(.) - D(.) \,] = -9.17$ nT
- Variance expérimentale $Var[\, D_4^*(.) - D(.) \,] = 1185.14$ nT^2

Si on supprime des statistiques la contribution de la trentaine de points explosifs, on obtient une variance expérimentale proche de celle de $D_3^*(.)$, ce qui confirme notre analyse visuelle .

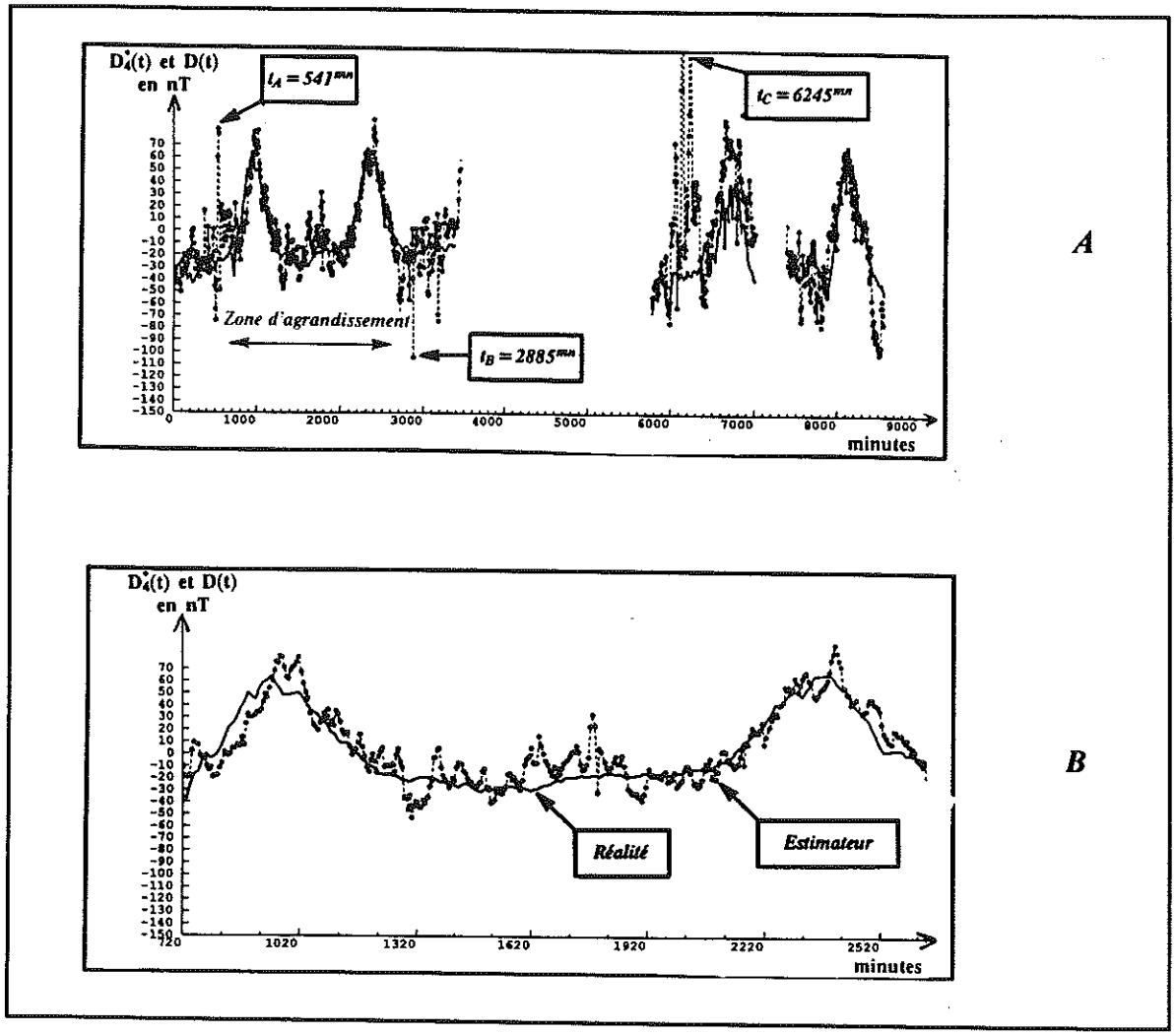


Figure 4.20 : Estimateur et réalité des perturbations (quatrième modèle)

Nous allons examiner dans le détail trois points (notés t_a, t_b et t_c sur la figure) qui sont explosifs . La figure 4.21 montre l'implantation dans le plan des échantillons qui sont intervenus dans l'estimation des points t_a, t_b et t_c .On note que les points contraints spatialement se situent tous dans la partie Sud-Est de la zone d'étude . Cette zone est constituée de forts gradients . On note de plus que la droite des points contraints temporellement étend son échantillonnage jusqu'à l'extrême Nord de la figure.

La figure 4.22 montre l'implantation dans le temps et par rapport à D(t) des échantillons qui sont intervenus dans l'estimation des points t_a, t_b et t_c . On note d'une part que globalement, cet échantillonnage est fait avec un pas supérieur à 12 h, et d'autre part que ces points se situent pratiquement tous dans la zone d'inactivité des perturbations (la nuit) et de ce fait n'échantillonnent aucune "tendance" temporelle .

Les tableaux 4.1 contiennent les valeurs de z(x,t) associées aux échantillons prélevés et les pondérateurs de Krigeage obtenus pour les trois points t_a, t_b et t_c . On note que les valeurs du magnétisme sont très hétérogènes au sein de l'ensemble Ψ , et qu'ils lui correspondent des pondérateurs non négligeables, de sorte que ces valeurs interviennent dans la combinaison linéaire constitutive de l'estimateur . On ajoutera à cela que les pondérateurs sont en moyenne 10 fois supérieurs à ceux obtenus en d'autres endroits .

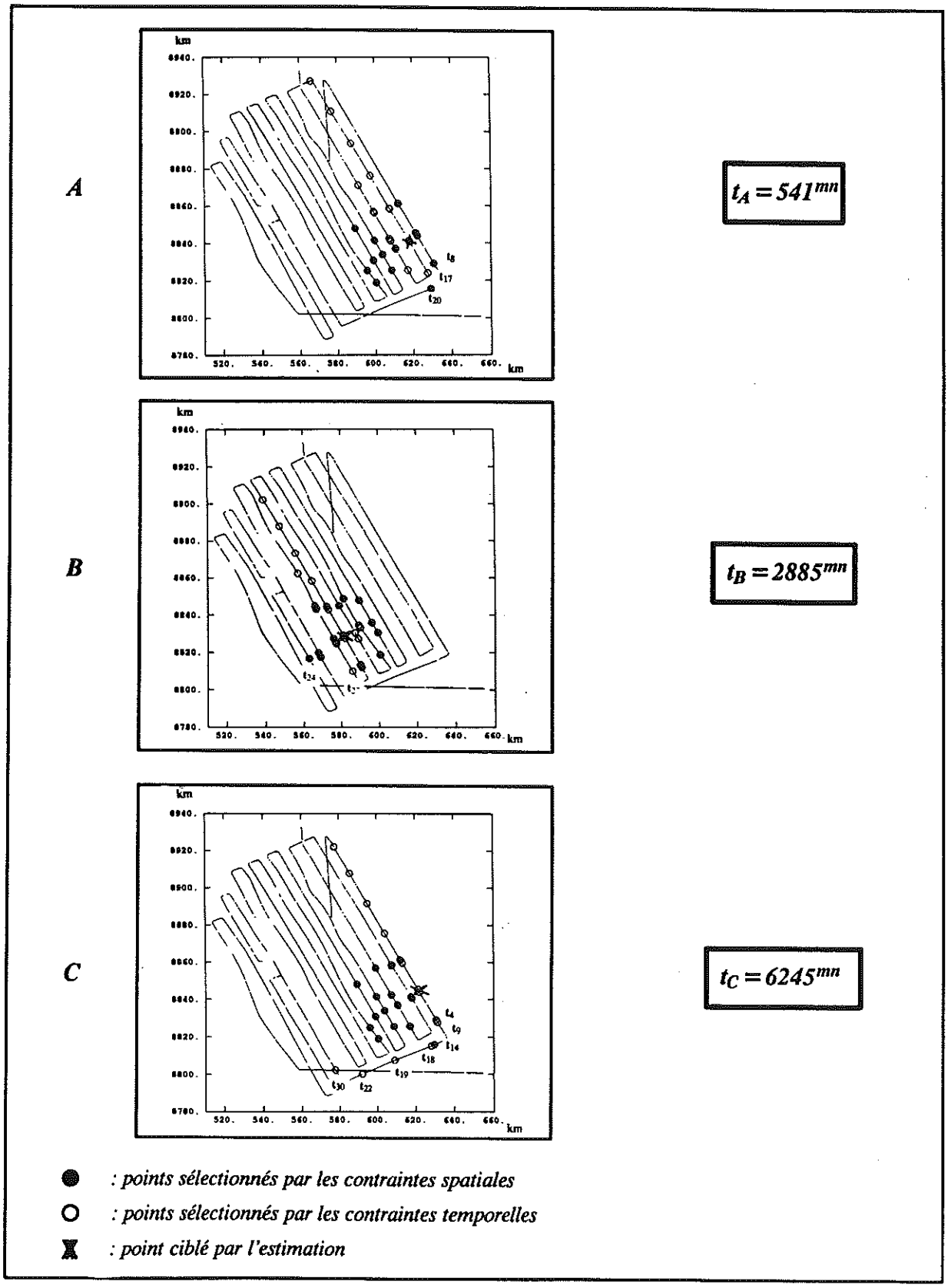


Figure 4.21 : Echantillonnage spatial correspondant aux points "explosifs"(quatrième modèle)

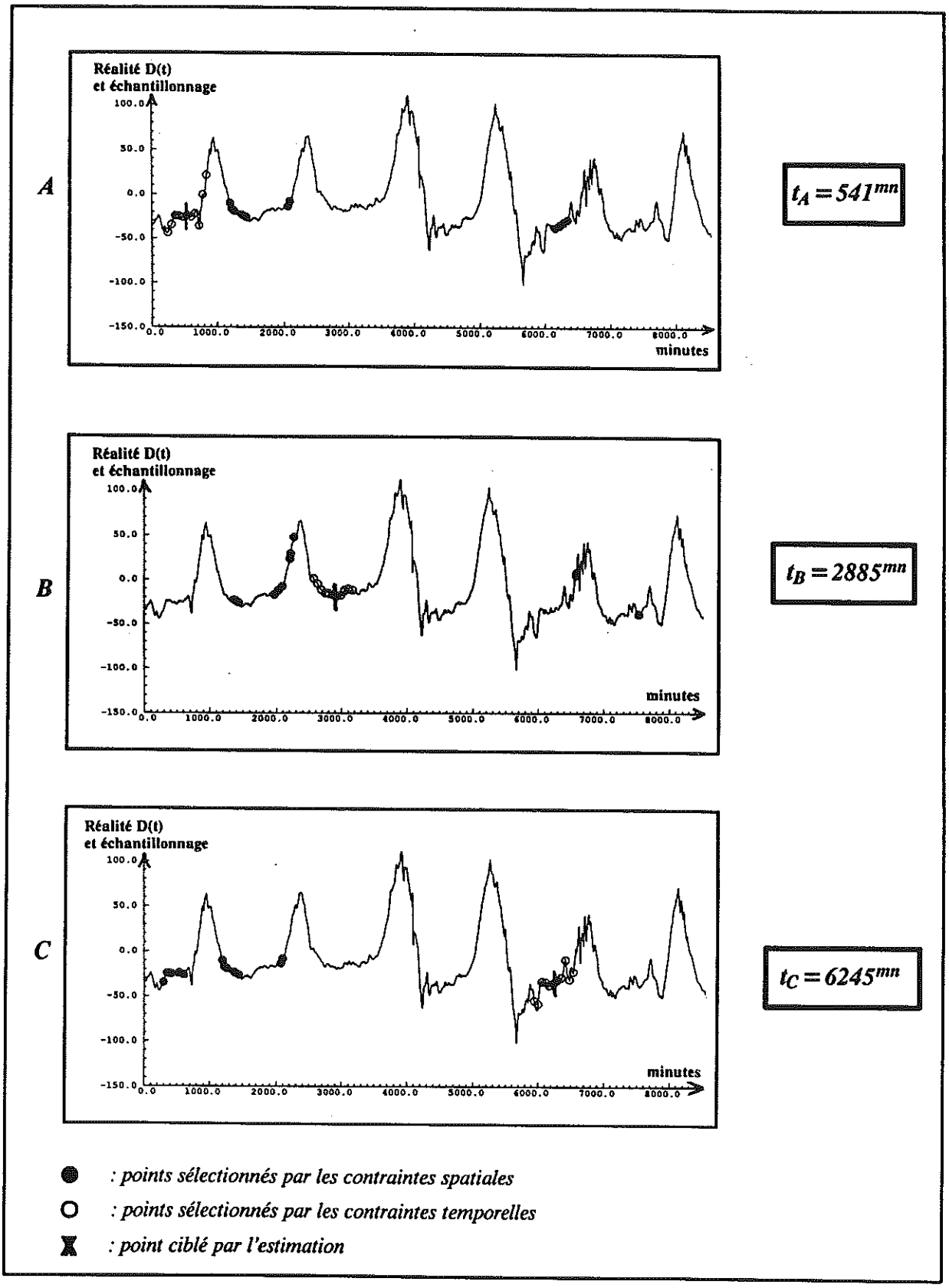


Figure 4.22 : Echantillonnage temporel correspondant aux points "explosifs"

$t_A = 541^{mn}$

rang	t_i	$z(t_i)$	$\lambda_x(t_i)$	
1	841	-4.3	-0.1536	○
2	781	-13.1	-0.0530	○
3	721	-25.4	0.0179	○
4	661	45.1	0.2282	○
5	241	-4.5	-0.2463	○
6	1440	-60.2	0.0777	●
7	2100	-99.2	-0.3662	●
8	6362	-233.8	-0.3100	● → t_8
9	2073	51.0	0.4108	●
10	299	-95.1	-0.1587	●
11	301	-95.7	-0.1482	○
12	1380	22.9	-0.0140	●
13	1379	27.7	-0.0154	●
14	6179	-72.7	0.2202	●
15	601	-74.3	0.1160	○
16	600	-74.5	0.0730	●
17	481	-212.6	0.4769	○ → t_{17}
18	1260	4.4	-0.1155	●
19	1199	-42.4	-0.1161	●
20	6300	-170.0	-0.1975	● → t_{20}
21	421	-88.3	0.2480	○
22	1228	-3.7	-0.0934	●
23	356	-18.9	-0.0766	●
24	361	-6.3	-0.0133	○
25	377	29.2	-0.0280	●
26	6239	33.3	0.0631	●
27	6245	59.8	-0.0452	●
28	539	75.1	0.1049	●
29	541	72.4	0.1197	⧗
30	6240	39.1	-0.0057	●

● : *points sélectionnés par les contraintes spatiales*

○ : *points sélectionnés par les contraintes temporelles*

⧗ : *point ciblé par l'estimation*

Tableau 4.1 : Pondérateurs de krigeage et quantité de magnétisme associés aux points "explosifs"

$t_B = 2885^{mn}$

rang	t_i	$z(t_i)$	$\lambda_x(t_i)$	
1	2885	-41.0	0.1848	⧗
2	2879	-41.8	0.1750	●
3	3056	-62.1	-0.0407	●
4	2940	-74.7	0.0892	●
5	1440	-60.2	0.3525	●
6	2219	36.8	-0.0116	●
7	2230	23.0	-0.1605	●
8	2825	-108.9	0.1925	○
9	3060	-69.3	-0.0267	●
10	2945	-93.1	0.0360	○
11	2100	-99.2	0.0809	●
12	2040	-26.1	0.1127	●
13	2280	-30.2	-0.2836	●
14	3065	-68.2	-0.0434	○
15	3005	-51.7	0.0184	○
16	2035	-31.6	-0.0931	●
17	2819	-98.2	0.1318	●
18	6607	-55.9	0.0320	●
19	1398	-33.0	0.0172	●
20	1979	-84.2	-0.1483	●
21	6599	-71.7	0.1082	●
22	1379	27.7	-0.2632	●
23	3120	-77.4	-0.0750	●
24	7546	-123.0	-0.3150	● → t_{24}
25	3125	-66.6	0.0127	○
26	2765	10.5	-0.0566	○
27	3185	-131.7	0.1965	○ → t_{27}
28	2705	-10.1	-0.0925	○
29	2645	-41.1	-0.0474	○
30	2585	-7.9	-0.0826	○

● : *points sélectionnés par les contraintes spatiales*

○ : *points sélectionnés par les contraintes temporelles*

⧗ : *point ciblé par l'estimation*

Tableau 4.1 : Pondérateurs de krigeage et quantité de magnétisme associés aux points "explosifs" (suite)

$t_C = 6245^{mn}$

rang	t_i	$z(t_i)$	$\lambda_x(t_i)$	
1	6245	59.8	-0.0462	X
2	6239	33.3	-0.0600	●
3	541	72.4	0.0751	●
4	6300	-170.0	0.2299	● → t_4
5	299	-95.1	-0.3399	●
6	421	-88.3	0.4307	●
7	6185	-72.3	0.1174	○
8	539	75.1	0.0886	●
9	6305	-190.3	0.0349	○ → t_9
10	1440	-60.2	0.0246	●
11	1260	4.4	-0.1246	●
12	6179	-72.7	0.1334	●
13	377	29.2	-0.0199	●
14	6360	-233.8	0.0044	● → t_{14}
15	2100	-99.2	-0.2027	●
16	600	-74.5	0.0135	●
17	356	-18.9	-0.0733	●
18	6365	-233.8	-0.1109	○ → t_{18}
19	6425	-158.9	-0.0584	○ → t_{19}
20	6125	14.7	0.1216	○
21	1228	-3.7	-0.0839	●
22	6485	-132.1	0.0417	○ → t_{22}
23	6065	-86.6	0.0049	○
24	1199	-42.4	-0.0378	●
25	6005	-46.9	-0.0346	○
26	1379	27.7	-0.0650	●
27	5945	-73.7	-0.0820	○
28	1380	22.9	-0.0200	●
29	2075	55.4	0.5433	●
30	6545	-123	-0.3649	○ → t_{30}

● : *points sélectionnés par les contraintes spatiales*

○ : *points sélectionnés par les contraintes temporelles*

X : *point ciblé par l'estimation*

Tableau 4.1 : Pondérateurs de krigeage et quantité de magnétisme associés aux points "explosifs" (fin)

Ces explosions sont donc dues à :

- La combinaison d'un contraste d'échelle entre les points contraints dans le temps et ceux contraints dans l'espace, qui réduit la géométrie à une droite dont l'extrémité Nord est constituée de valeurs tempérées, et l'extrémité Sud de valeurs très fluctuantes et fortement négatives .
  Or, le système est conduit avec $k=1$, et on sait qu'une droite de point est une configuration interdite car elle rend les polynômes de l'espace linéairement dépendants sur les données.

- L'échantillonnage dans le temps qui, pour ces points explosifs, possède un pas inadapté à la période 12^h, et même sans doute à 24^h .

- Les valeurs qui sont échantillonnées dans le temps sont toutes sensiblement identiques et ne permettent donc pas un bon "calage" des sinusoïdes du système .

Le problème est de savoir laquelle de ces trois circonstances est la plus importante . Dans ce but, nous allons reconduire l'estimation en FAI-$\emptyset$.

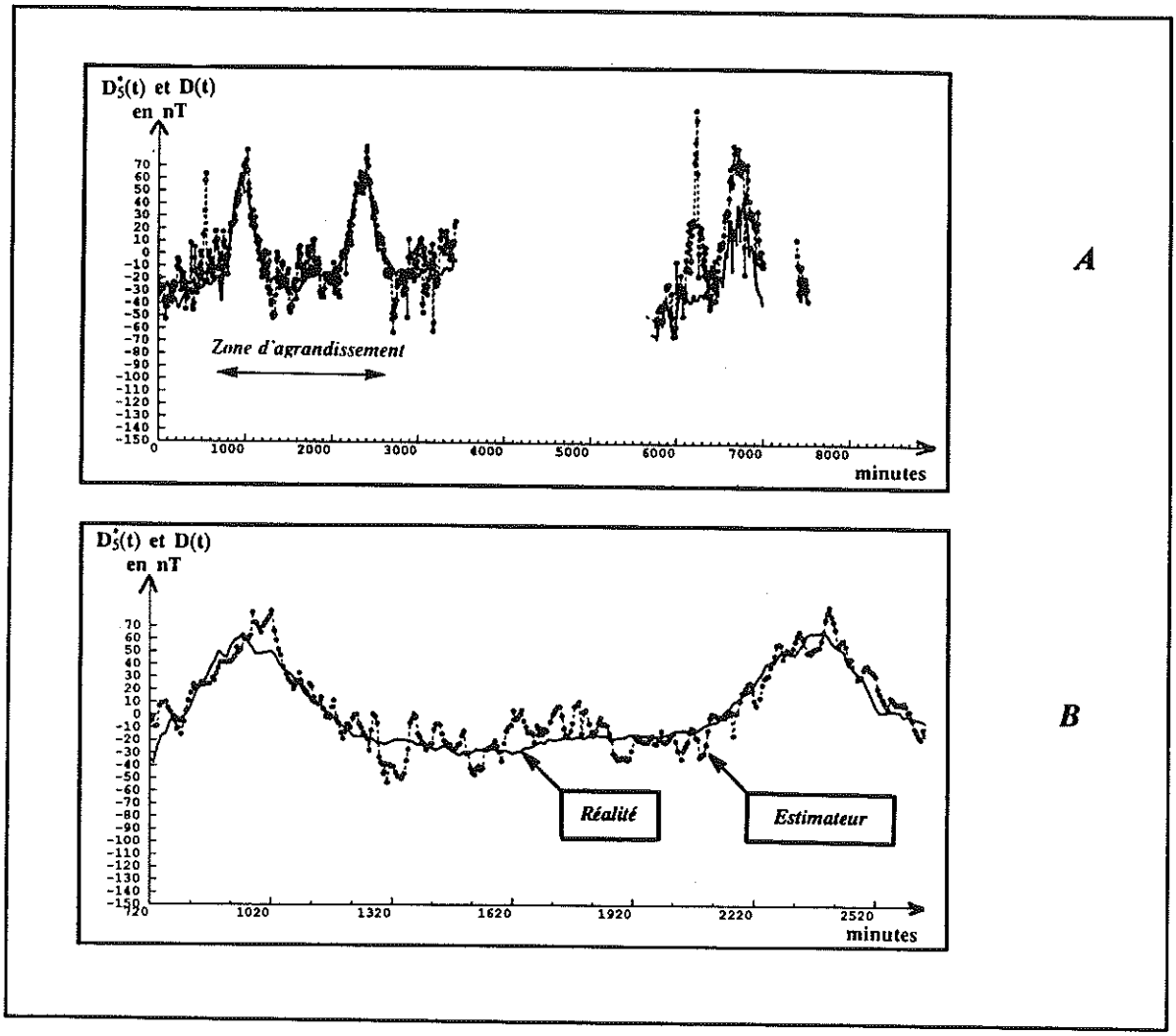


Figure 4.23 : Estimateur et réalité des perturbations (cinquième modèle)

4 .3 .7 Estimation avec le cinquième modèle

Le modèle utilisé est :

$$\left\{ \mathcal{V}_4^{x\,t} \; ; \; k_{prior} = \emptyset \; ; \; T_{prior} = 24^h, 12^h \; ; \; K^5(h) \right\}_{prior} \text{ avec } K^5(h) = a_1 \, |h| \, , \, a_1 = -3.672$$

On rappelle que ce modèle est intégralement arbitraire . Outre la CG utilisée, il ne diffère du précédent que par le degré de la FAI qui a été porté arbitrairement à $\emptyset$.

On note sur la figure 4.23 qui représente l'estimateur obtenu, que bon nombre des explosions précédentes ont disparu.
Les paramètres statistiques confirment le jugement visuel :

- Nombre de points estimés = 5395
- Erreur moyenne $E[\, D_5^*(.) - D(.) \,] = -10.55$ nT
- Variance expérimentale $Var[\, D_5^*(.) - D(.) \,] = 497.28$ nT^2

On peut donc que ces instabilités étaient essentiellement dues au fait que la géométrie du type $\mathcal{V}_4^{x\,t}$ est équivalente à une droite, ce qui est incompatible avec une valeur non nulle du paramètre k .

4 .3 .8 Conclusions sur Seaperc

Pour les praticiens contactés, ce filtrage est un succès .
Les figures 4.24 et 4.25 montrent comment se situe notre estimateur par rapport aux cartes vraies et non filtrées .

Il ressort de ce travail qu'une fois de plus, c'est la géométrie du type $\mathcal{V}_3^x$ qui s'est avérée la plus adaptée au système de KT .

On ne manquera pas de s'étonner des variances expérimentales obtenues qui sont du même ordre de grandeur que celles des expériences mono-dimensionnelles citées (signal d'entrée : les perturbations temporelles et non le signal spatio-temporel) .

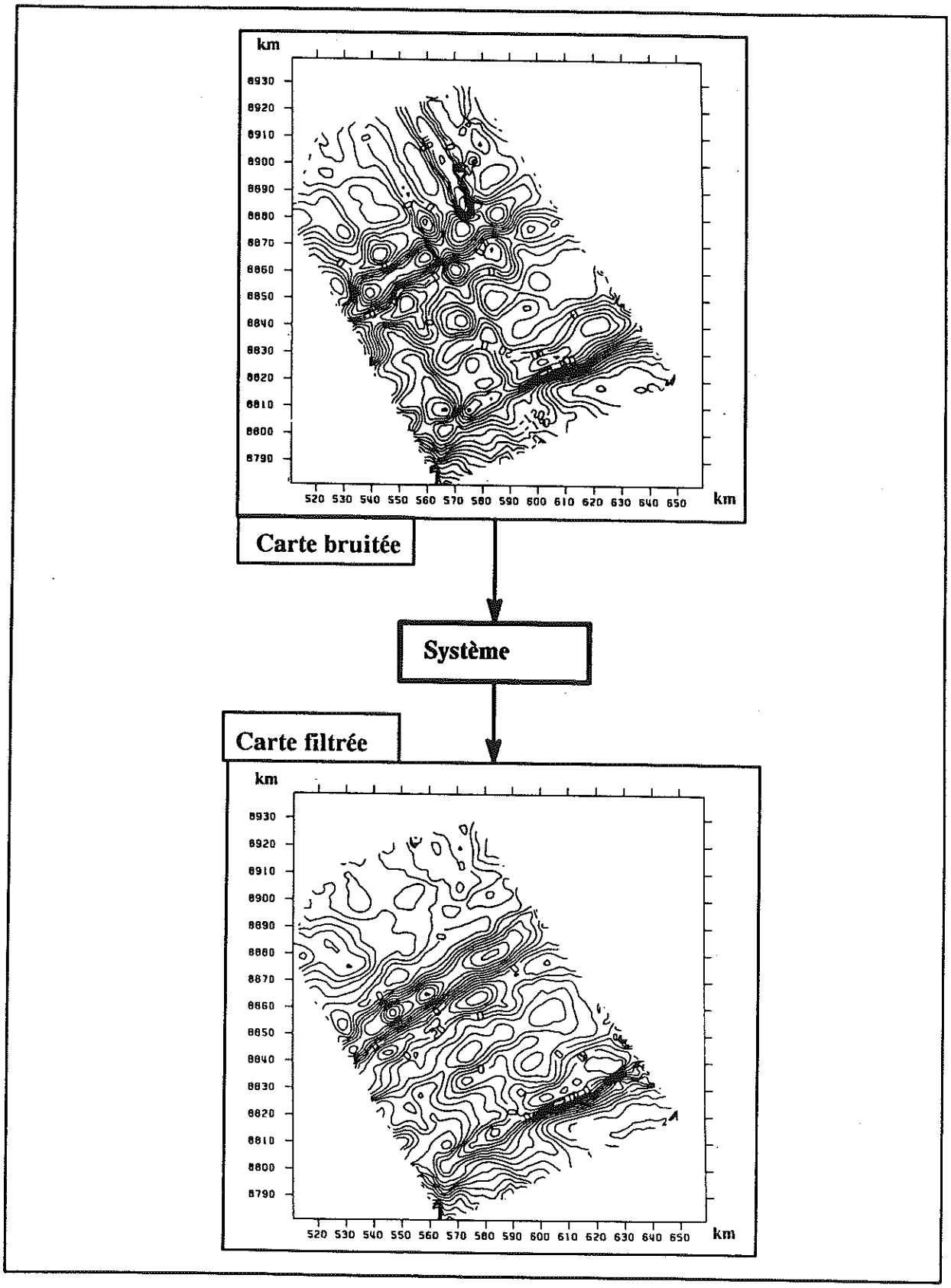


Figure 4.24 : Le chemin parcouru grâce au filtrage

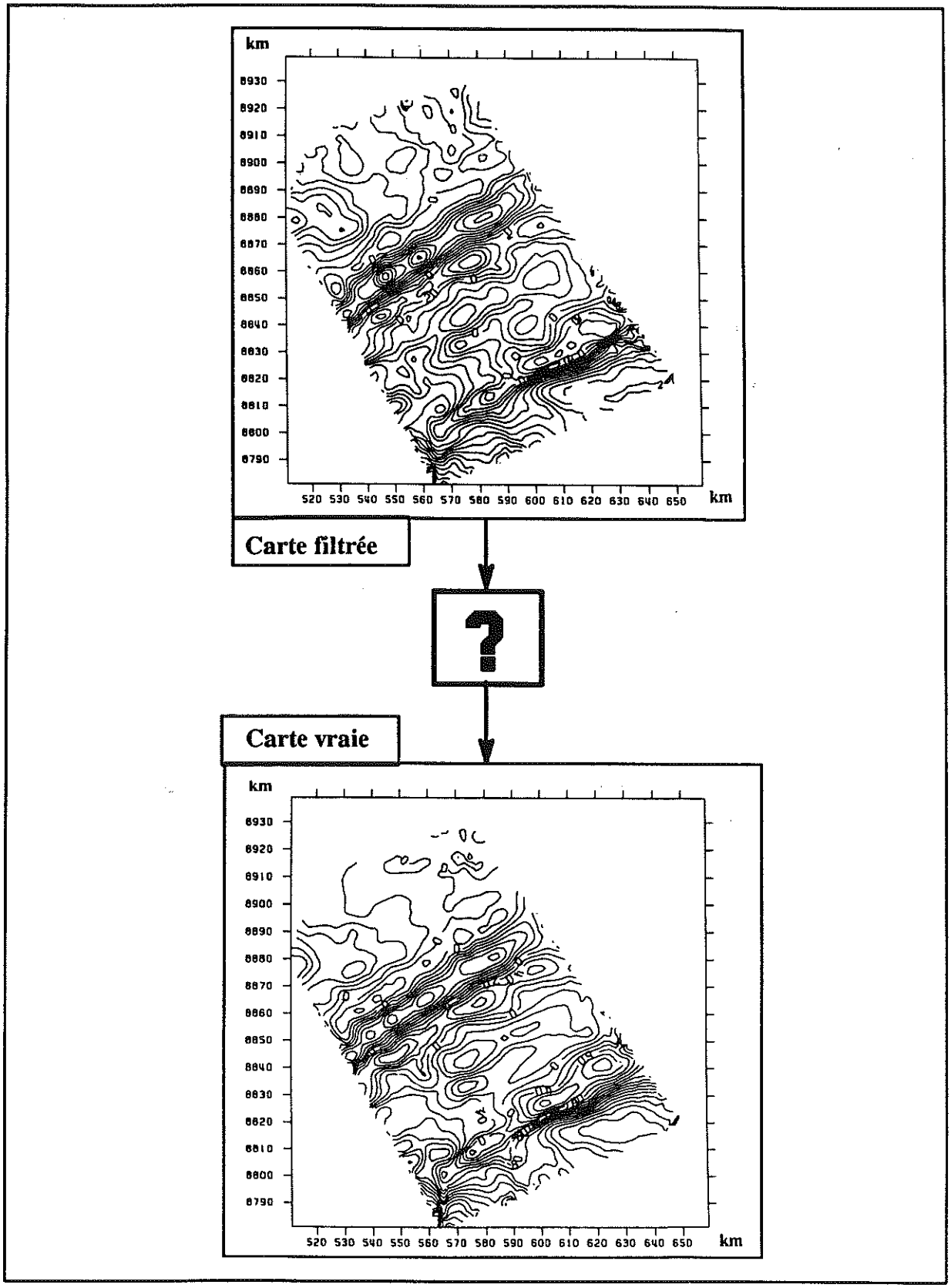


Figure 4.25 : Le chemin qu'il reste à parcourir ...

4 .4 Filtrage des données de la campagne MOHNS

Ce dernier cas d'étude se caractérise par :

- Des approches globale (champ complet) et locale .
- Des fluctuations temporelles qui ne sont pas périodiques .
- Le fait que nous ne disposons que de 3 jours de mesure des perturbations sur zone, et qui plus est sans connaître avec exactitude l'origine des temps de celle-ci . Ceci explique que des analyses comparées entre variogrammes de l'estimateur et de la réalité seront faites .

4 .4 .1 Présentation des données

Nous rappelons succinctement la carte du magnétisme lorsqu'il n'est pas filtré des perturbations sur la zone globale (figure 4.26) puis locale (figure 4.27) et l'allure des perturbations près de la zone d'étude (figure 4.28). Comme D(t) n'est connu que sur 3 jours, cette série temporelle n'a pas pu permettre une correction de référence de z(x,t) .

Ce cas d'étude contiendra également une analyse spectrale des estimateurs obtenus. C'est pourquoi nous rappelons l'allure du spectre d'énergie associé à D(t) (série Mohns, figure 4.29) et le tableau de leurs valeurs (tableau 4.2) .

4 .4 .2 Les modèles disponibles

Les modèles disponibles sont aux nombre de 4 :

- $\left\{ \varphi_2^x \; ; \; k = 1 \; ; \; T = 24^h \; ; \; K^1(h) \right\}^{\text{global}}_{\text{prior}}$ avec $K^1(h) = a_1 \, |h|$ $\Big\} \longrightarrow D_1^*(t)$

- $\left\{ \varphi_2^x ; k = 2 ; T_{prior} = 24^h ; K^2(h) = K^{prior}(h) \right\}^{\text{local}}$ avec $K^2(h) = a_\emptyset \, \text{pépite}(h) + a_3 \, |h|^3$ $\Big\} \longrightarrow D_2^*(t)$

- $\left\{ \varphi_3^x ; k = 1 ; T = 24^h ; K^3(h) = K^{prior}(h) \right\}^{\text{global}}_{\text{prior}}$ avec $K^3(h) = a_1 \, |h|$ $\Big\} \longrightarrow D_3^*(t)$

- $\left\{ \varphi_3^x \; ; \; k = 2 \; ; \; T_{prior} = 24^h, 12^h \; ; \; K^4(h) \right\}^{\text{local}}$ avec $K^4(h) = a_1 \, |h|$ $\Big\} \longrightarrow D_4^*(t)$

On rappelle que les exposants "local" et "global" désignent la zone concernée par les modèles, et que seuls les modèles locaux ne sont pas intégralement arbitraires .

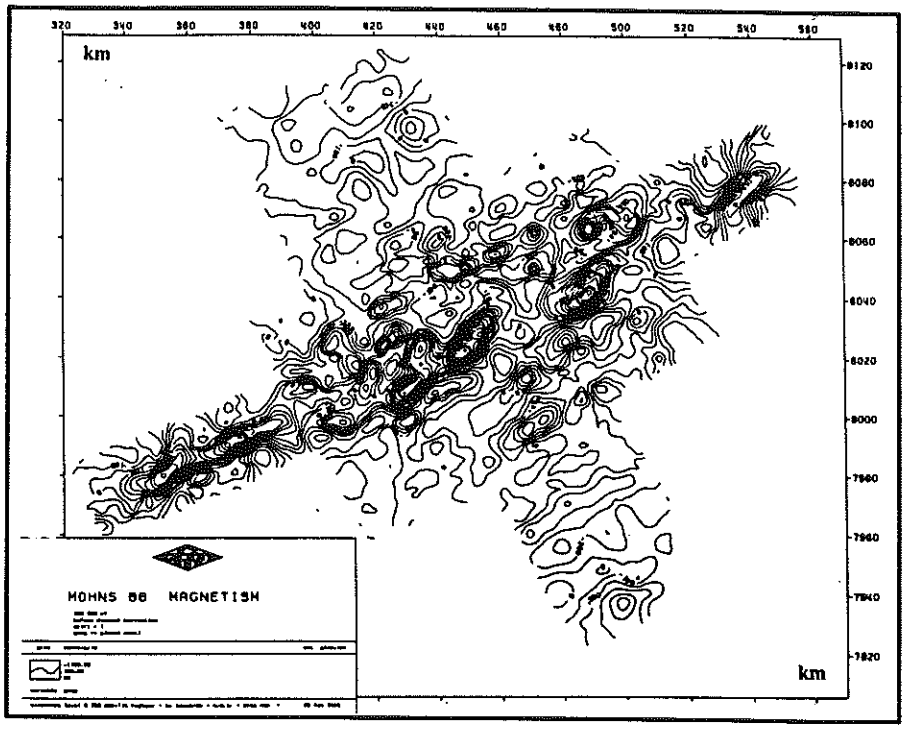


Figure 4.26 : Magnétisme spatio-temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes (zone globale)

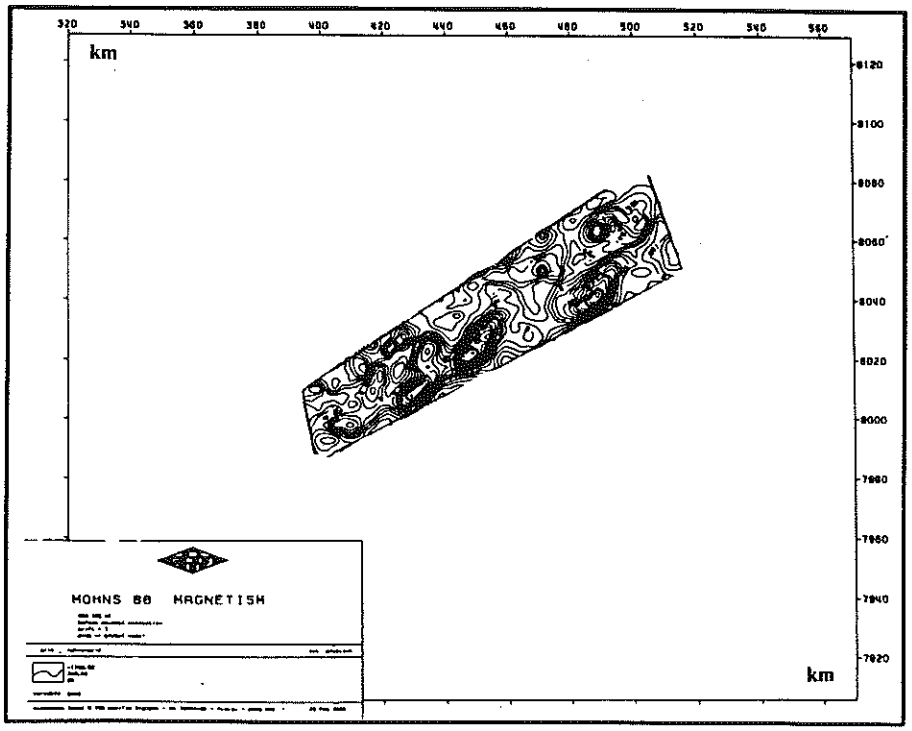


Figure 4.27 : Magnétisme spatio-temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes (zone locale)

Fréquence ω_i	Périodes (minutes) T_i	Energie $\xi = TFD(C_D^*)$	
0.0000000	0.00	6442.41	→ *Constante de Fourier*
0.0021817	2880.00	1783.21	→ *Fréquence émergente*
0.0043633	1440.00	1678.35	→ *Fréquence émergente*
0.0065450	960.00	499.53	→ *Fréquence émergente*
0.0087266	720.00	1065.92	→ *Fréquence émergente*
0.0109063	576.00	116.96	
0.0130900	480.00	427.62	→ *Fréquence émergente*
0.0174533	360.00	128.56	
0.0196350	320.00	106.00	
0.0218166	288.00	49.81	
0.0239983	261.82	127.79	
0.0261799	240.00	19.49	
0.0283616	221.54	35.30	
0.0305433	205.71	16.73	
0.0327249	192.00	103.93	
0.0349066	180.00	36.22	
0.0370882	169.41	40.86	
0.0414516	151.58	69.27	
0.0436332	144.00	21.36	
0.0545415	115.20	14.19	

Tableau 4.2 : Energies aux fréquences associées aux perturbations de l'enregistreur immergé en bordure de la zone d'étude

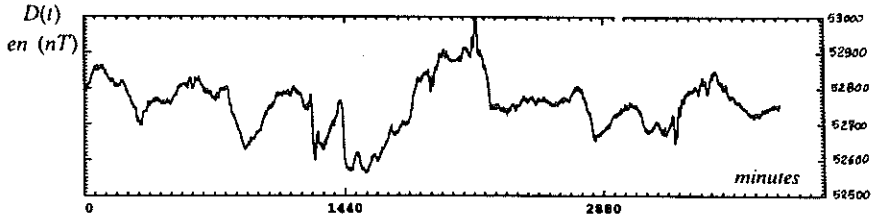


Figure 4.28 : Evolutions dans le temps du champ magnétique total donné par l'enregistreur immergé

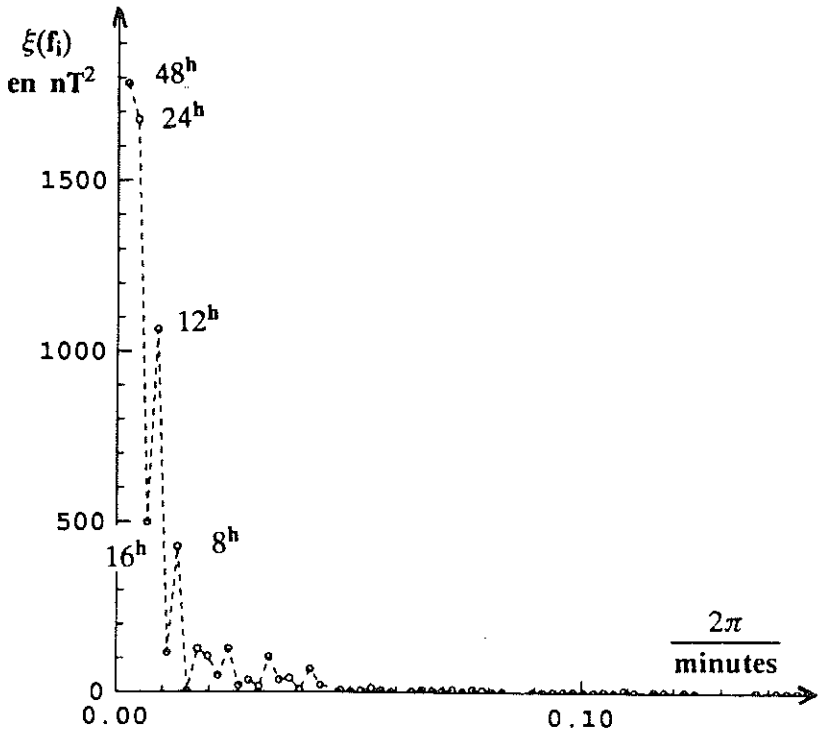


Figure 4.29 : Spectre d'énergie des perturbations

4 .4 .3 Estimation avec le premier modèle

Le modèle utilisé est :

$$\left\{ \mathcal{V}_2^x \ ; \ k = 1 \ ; \ T = 24^h \ ; \ K^1(h) \right\}_{\text{prior}}^{\text{global}} \text{ avec } K^1(h) = a_1 \ |h| \text{ et } a_1 = -4990$$

On rappelle que ce modèle est intégralement "a-priori"

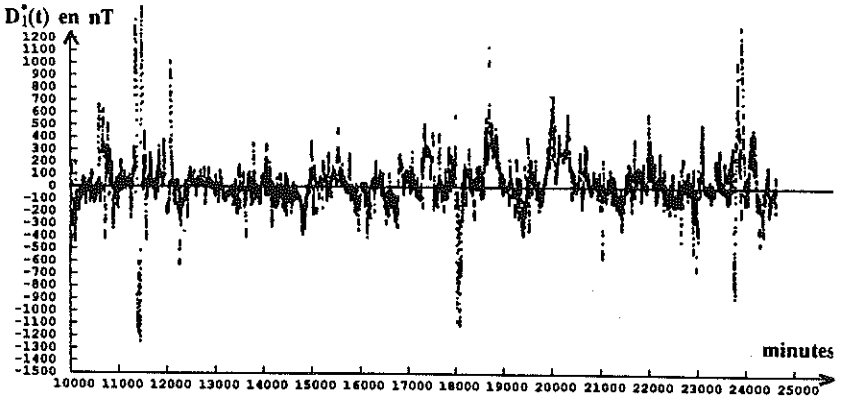


Figure 4.30 : Perturbations estimées (premier modèle)

L'estimateur obtenu (figure 4.30) présente les propriétés suivantes :

- Il est fortement bruité .
- Il présente de nombreuses et importantes explosions .
- La tendance périodique n'est pas manifeste .

La figure 4.31 et le tableau 4.3 représentent le spectre d'énergie associé à cet estimateur .

Fréquence ω_i	Périodes (minutes) T_i	Energie $\xi = TFD(C_D^*)$	
0.0000000	0.00	49945.30	→ *Constante de Fourier*
0.0021817	2880.00	1160.92	→ *48 h*
0.0043633	1440.00	7387.75	→ *24 h*
0.0130900	480.00	2775.75	→ *8 h*
0.0349066	180.00	2053.91	→ *3 h*
0.0523599	120.00	2090.83	→ *2 h*

Tableau 4.3 : Energies aux fréquences de l'estimateur (premier modèle)

Nos remarques :

- L'énergie à 24 h est considérable et contient à elle seule la quasi totalité des énergies en jeu . Cette énergie de 7388 nT^2 pour 24 h est sans commune mesure avec celle contenue dans D(.) pour la même période et qui vaut 1678 nT (tableau 4.2) .
  Elle est due sans doute à un phénomène de diffusion : la TFD ayant été conduite sur un diamètre de 48 h (afin de pouvoir comparer ses résultats à ceux de D(.)), l'énergie

de l'estimateur (et notamment l'énergie des explosions) se répartit alors sur les seules fréquences données par l'analyseur .

- L'émergence d'amplitudes à 8 h 3 h et 2 h .
  On associera ces amplitudes aux instabilités numérique.

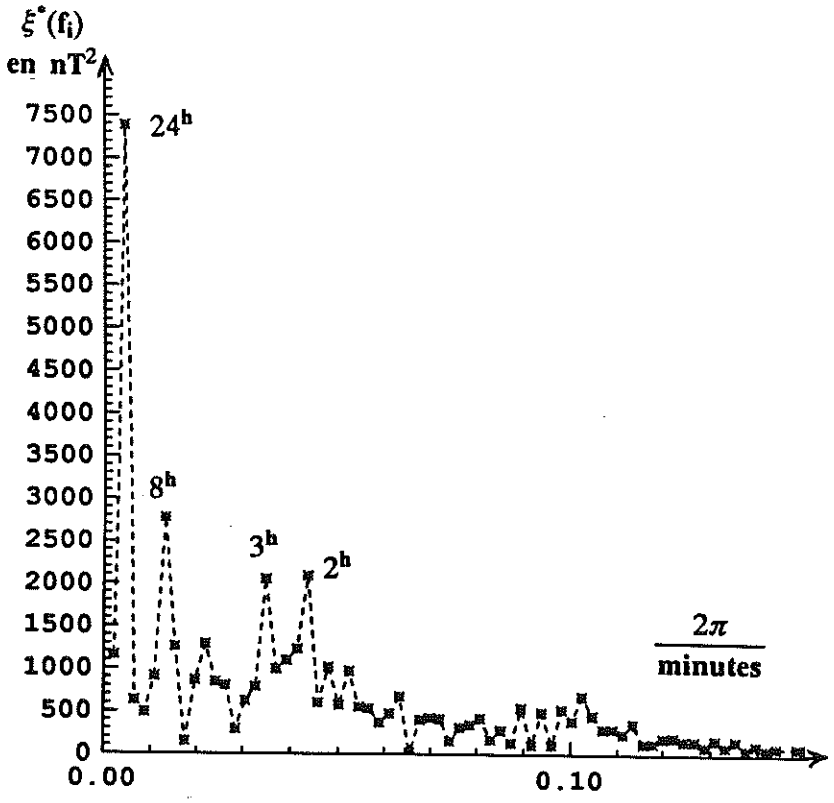


Figure 4.31 : Spectre d'énergie de l'estimateur (premier modèle)

Nous présentons sur la figure 4.32 six jours du variogramme de ce premier estimateur .

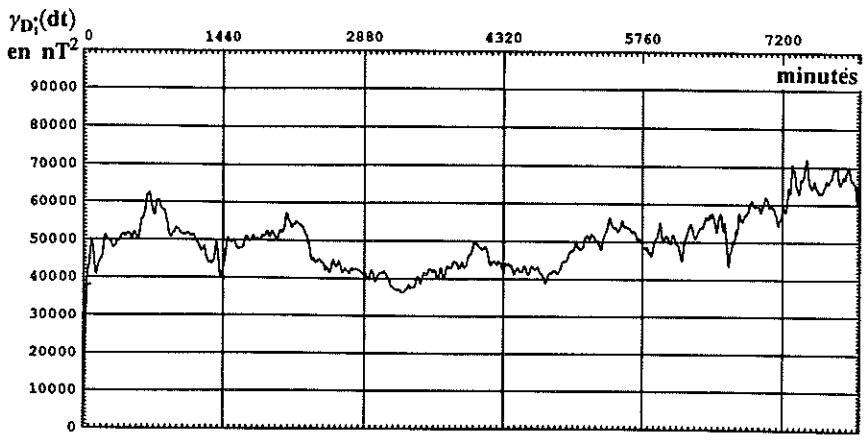


Figure 4.32 : Variogramme de l'estimateur (premier modèle)

Il paraît difficile de reconnaître là le variogramme d'un PTP . Cette allure non périodique confirme que le spectre précédent calculé sur 3 jours n'est pas représentatif des véritables énergies en jeu, et notamment celle associée à 24 ^{h} .

Nous présentons sur la figure 4.33, un agrandissement du variogramme de l'estimateur, comparativement à celui obtenu à l'aide des 3 jours de mesure de D(.) .

Ces variogrammes n'ont aucun lien de parenté et sont dans un rapport de variances de 1 à 10,

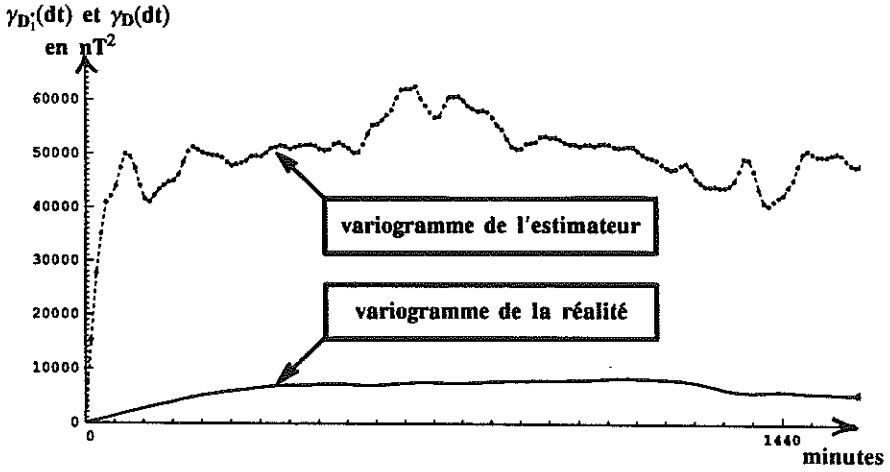


Figure 4.33 : Variogrammes de la réalité et de l'estimateur (premier modèle)

La carte obtenue à l'aide de ce premier estimateur (figure 4.34) est également bruitée et ne constitue pas une amélioration par rapport à la carte sans filtrage .

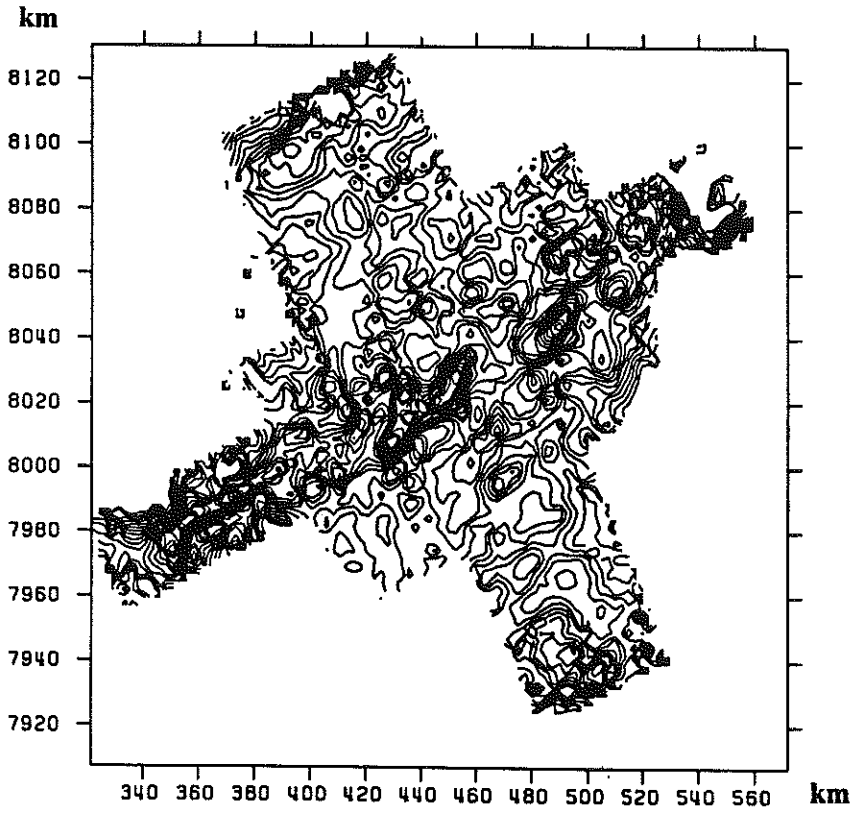


Figure 4.34 : Anomalies spatiales estimées (premier modèle)

4 .4 .4 Estimation avec le deuxième modèle

Le modèle utilisé est :

$\{ \mathcal{V}_2^x ; k = 2 ; T_{prior} = 24^h ; K^2(h) = K^{prior}(h) \}^{local}$

avec $K^2(h) = a_\beta$ pépite(h) + $a_3 |h|^3$, $a_\beta = 1\,151.$, $a_3 = 179.9$,

Dans ce modèle, l'introduction d'une fréquence est arbitraire puisque celle-ci n'a pas pu être inférée lors de l'analyse et la modélisation. De même, la CG utilisée n'était pas celle préconisée par la machine.

La figure 4.35 représente l'estimateur temporel obtenu. Les lacunes de l'estimation sont dues au fait que les estimations n'ont été conduites que sur les temps impliqués par la zone locale.

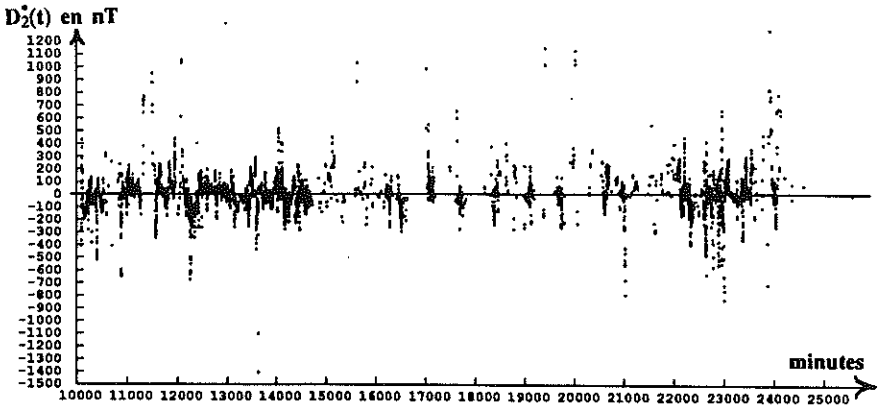


Figure 4.35 : Perturbations estimées (deuxième modèle)

Comparativement à l'estimateur précédent, on peut remarquer que :

- Les explosions sont synchrones .
  Comme la géométrie $\mathcal{V}_2^x$ est commune à ces deux estimateurs, ce synchronisme implique la géométrie une fois de plus .
- Ces explosions sont d'amplitude plus importantes que précédemment.
  La raison en est sans doute que ce deuxième modèle possède k = 2 pour degré de FAI.
- Hormis les zones explosives, les évolutions des estimateurs sont comparables .
  Le passage du global au local a donc peu influencé l'estimation .

Comme la zone couverte par l'enregistrement D(t) ne correspond pas aux temps des mesures sur la zone locale, nous ne présentons pas de spectres d'amplitude pour cet estimateur local .

La carte de l'estimateur spatial (figure 4.36) n'apporte rien de nouveau par rapport à la même zone de $Y_1^*(x)$.

Conclusion sur le deuxième modèle :

- Le passage d'un modèle global à un modèle local n'a pas influencé l'estimation .
- La géométrie $\mathcal{V}_2^x$ est une fois de plus inadaptée à ce modèle .
- Ce filtrage n'en est pas un puisqu'il engendre une carte plus bruitée encore que celle obtenue sans filtrage .

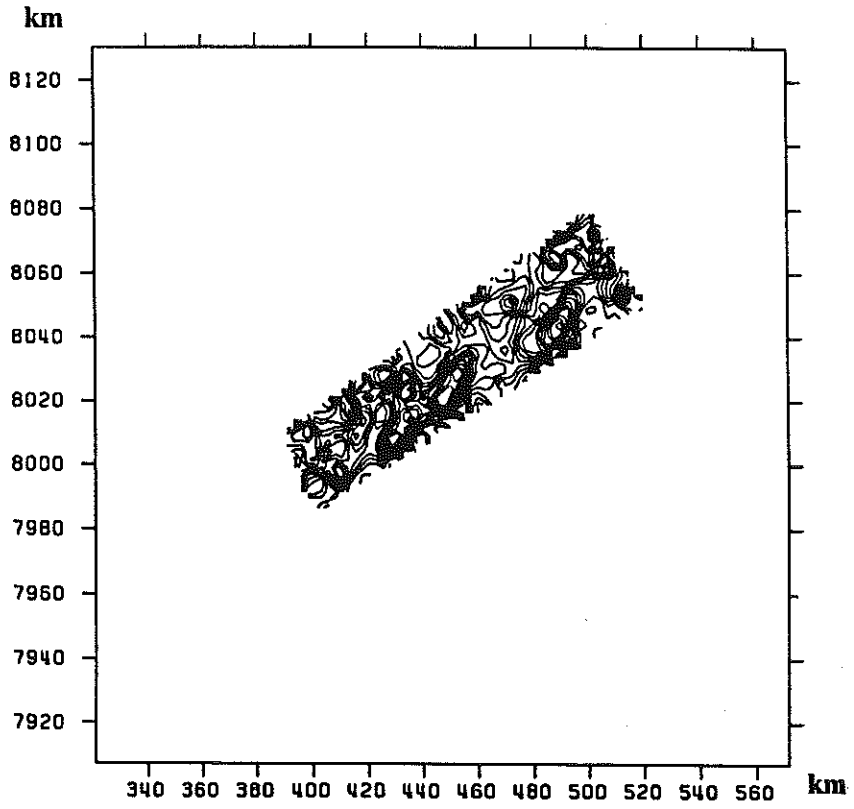


Figure 4.36 : Anomalies spatiales estimées (deuxième modèle)

4 .4 .5 Estimation avec le troisième modèle

Le modèle utilisé est :

$$\left\{ \nabla_3^{\chi} \; ; \; k = 1 \; ; \; T = 24^h \; ; \; K^3(h) = K^{prior}(h) \right\}_{prior}^{global} \text{ avec } K^3(h) = a_1 \, |h| \text{ et } a_1 = -14\,870$$

On rappelle que ce modèle est intégralement "a-priori"

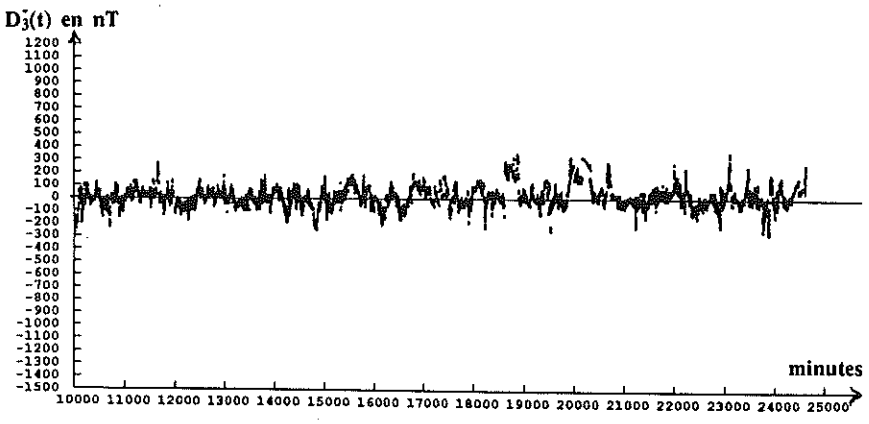


Figure 4.37 : Perturbations estimées (troisième modèle)

Le nouvel estimateur (figure 4.37) est beaucoup plus stable que les précédents . Mais cette stabilité est-elle un critère de justesse puisque ce qui subsiste ressemble plus à un bruit blanc qu'autre chose ?

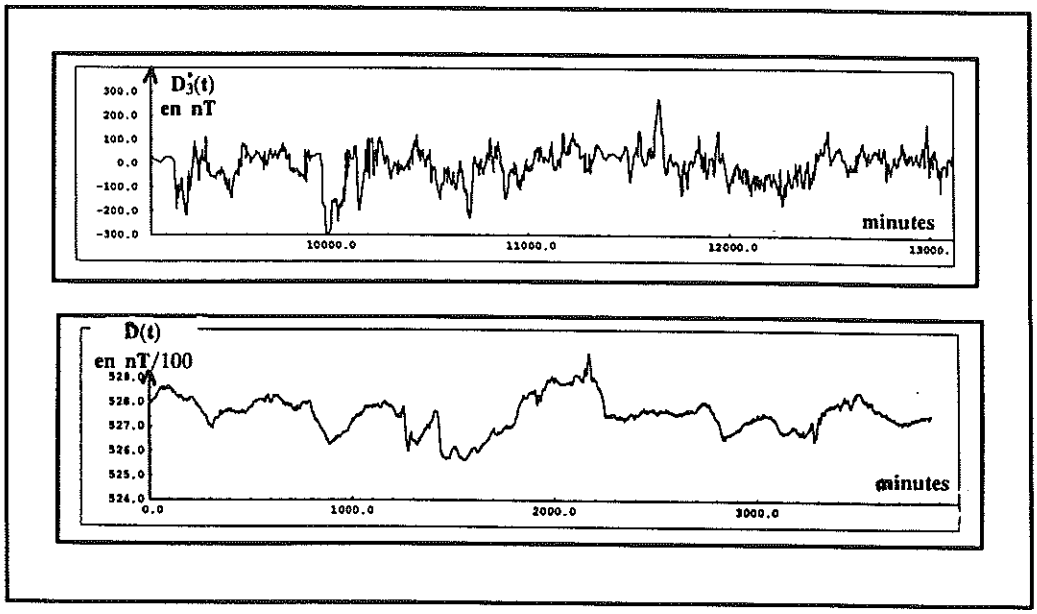


Figure 4.38 : Estimateur et réalité des perturbations (troisième modèle)

Nous présentons sur la figure 4.38 les évolutions comparées de l'estimateur et de la réalité . On peut y voir certaines similitudes, mais elles sont loin d'être immédiates . Ces éventuelles similitudes sont d'autant plus floues que l'origine des temps de D(t) n'est connue qu'à 200 ^{mn} près .

Le tableau 4.4 et la figure 4.39 présentent les énergies en jeu dans ce troisième estimateur .

Fréquence ω_i	Périodes (minutes) T_i	Energie $\xi = TFD(C_D^*)$	
0.0000000	0.00	6933.55	
0.0021817	2880.00	506.89	→ *48 h*
0.0043633	1440.00	1587.57	→ *24 h*
0.0087266	720.00	628.51	→ *12 h*
0.0109083	576.00	414.85	→ *9.6 h*
0.0130900	480.00	266.03	→ *8 h*
0.0218166	288.00	287.31	→ *4.8 h*

Tableau 4.4 : Energies aux fréquences de l'estimateur (troisième modèle)

On note que l'allure de ce spectre est comparable à celle du spectre de D(t) . Il en est de même de l'énergie associée à la période 24 ^{h} qui est de 1588 nT^2 alors que l'énergie réelle est , pour la même zone temporelle, de 1678 nT^2 .

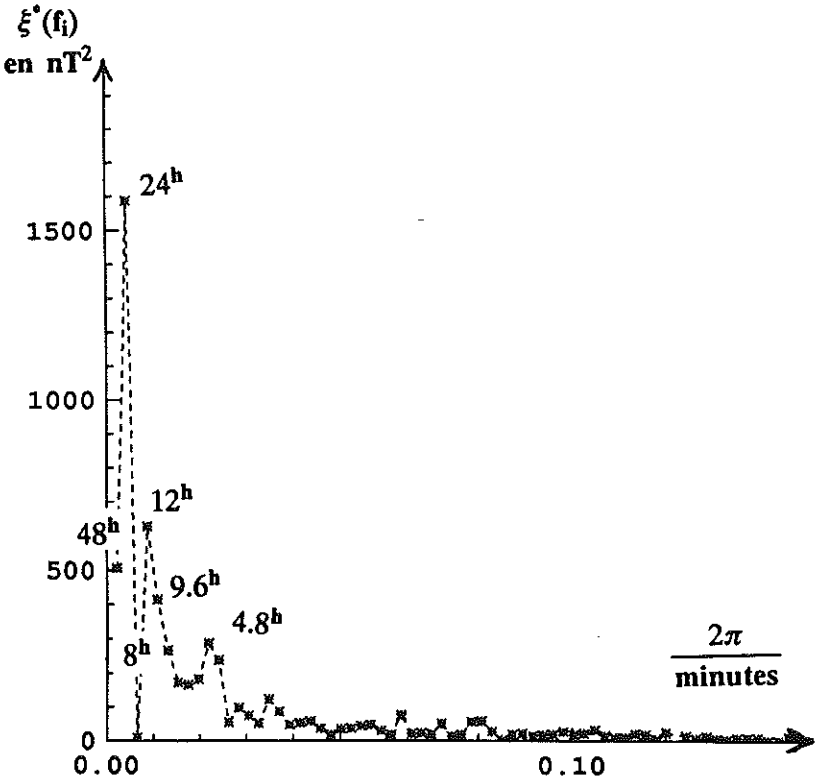


Figure 4.39 : Spectre d'énergie de l'estimateur (troisième modèle)

Le variogramme de l'estimateur (figure 4.40) est caractéristique, par ses oscillations, de celui d'un PTP .

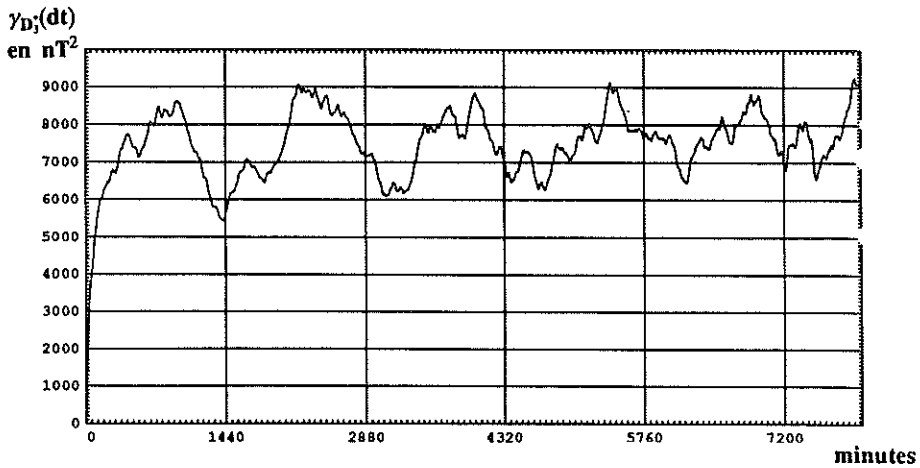


Figure 4.40 : Variogramme sur 5 jours de l'estimateur (troisième modèle)

Lorsqu'on le compare à celui de D(t) (figure 4.41), on constate également des similitudes. Mais ces similitudes possèdent-elles un sens lorsqu'on sait que :

- Le variogramme d'un estimateur n'a aucune raison d'être comparable à celui de la réalité (effet de la convolution) .
- Le variogramme de l'estimateur a été calculé sur 233 h, alors que celui de D(.) sur 56h seulement .

Nous ne le pensons pas .

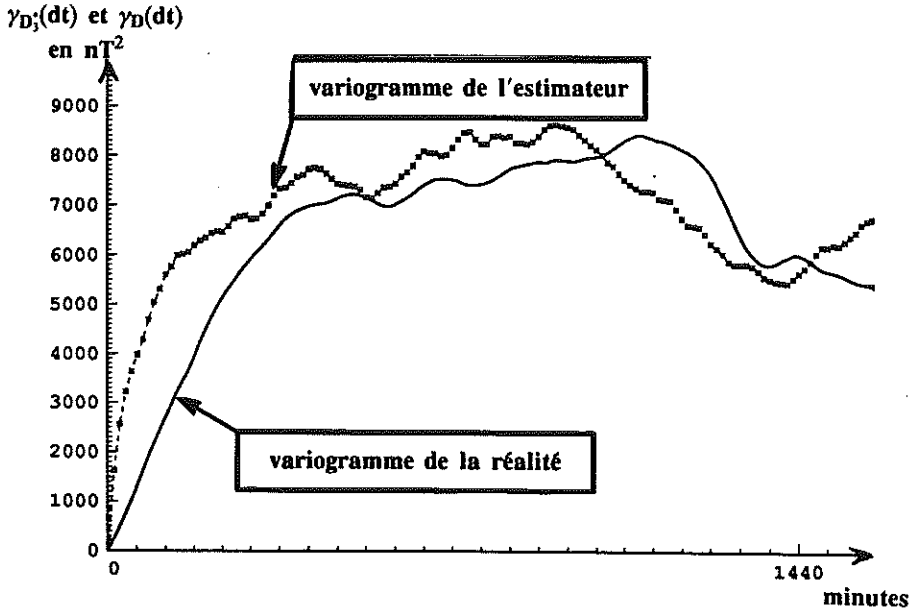


Figure 4.41 : Variogrammes de la réalité et de l'estimateur (troisième modèle)

La carte obtenue de l'estimateur spatial (figure 4.42) est à l'image de celle de l'estimateur temporel : moins bruitée . Certains des Géophysiciens interrogés voient en cette carte une amélioration, d'autres non . Que dire de plus ?

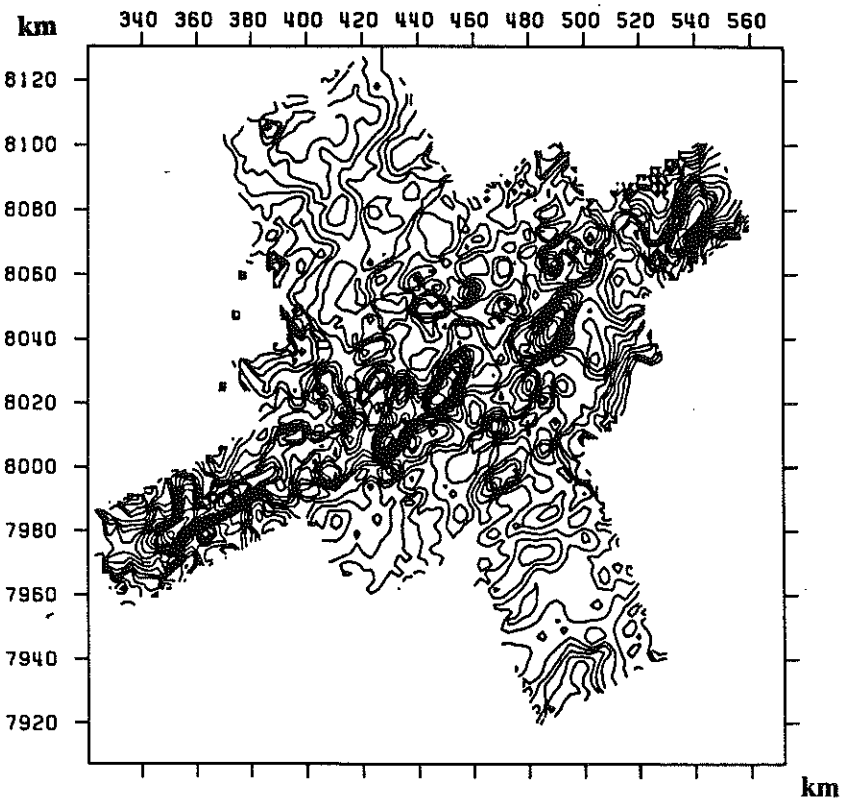


Figure 4.42 : Anomalies spatiales estimées (troisième modèle)

4 .4 .6 Estimation avec le quatrième modèle

Le modèle est :

$$\left\{ \nabla^x_3 \; ; \; k = 2 \; ; \; T_{prior} = 24^h, 12^h \; ; \; K^4(h) \right\}^{local} \text{ avec } K^4_{mohns}(h) = a_1 \, |h| \, , a_1 = -111\,90$$

Dans ce modèle, l'introduction de fréquences est arbitraire puisque celles-ci n'ont pas pu être inférées lors de l'analyse et la modélisation .

La courbe 4.43 permet de juger l'estimateur obtenu. On note que l'introduction de la période 12 ^{h} dans le système a entraîné de nouvelles explosions, et ceci malgré la géométrie utilisée . On se retrouve donc dans la même circonstance que les deux premiers modèles, et il est inutile de poursuivre l'analyse .

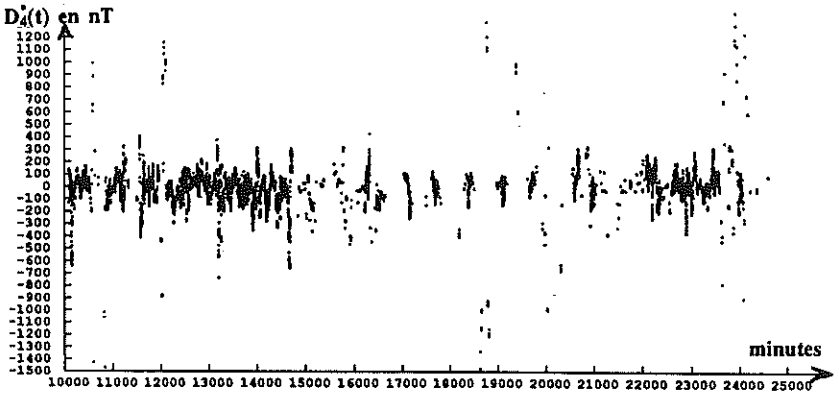


Figure 4.43 : Perturbations estimées (quatrième modèle)

On présente en 4.44 la carte des anomalies spatiales obtenues. Aucune remarque particulière à faire sur cet estimateur une fois de plus explosif.

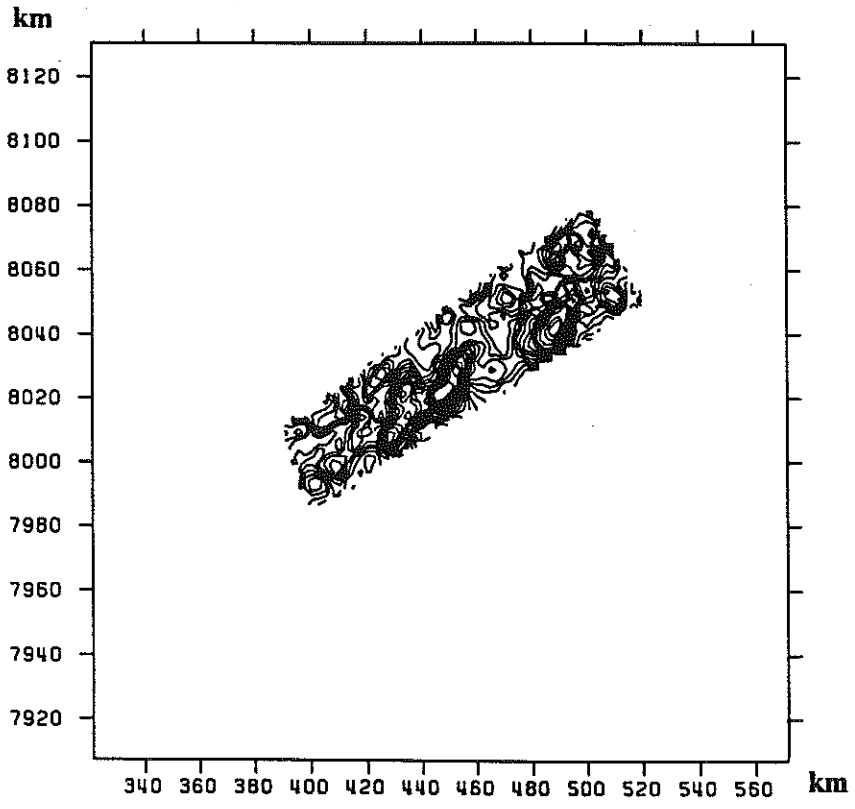


Figure 4.44 : Anomalies spatiales estimées (quatrième modèle)

4 .4 .7 Modification du pas d'échantillonnage

Dans une dernière tentative pour obtenir un filtrage efficace, nous avons effectué une nouvelle estimation, en utilisant le même modèle que celui de $D_3^*(t)$ (qui semble être le moins mauvais estimateur), mais en modifiant le pas d'échantillonnage et le nombre de points de la géométrie . Les modifications sont :

- Le pas est passé de 1 ^{h} à 30 ^{mn} .
  Ce pas est moyen et concerne la longueur des "pseudo-profils" .

- La population du voisinage est passée de 20 à 40 points .
  Ceci dans le but de préserver le diamètre du voisinage

La figure 4.45 permet de constater que l'estimateur temporel est devenu plus régulier encore et s'apparente de plus en plus à un bruit blanc .

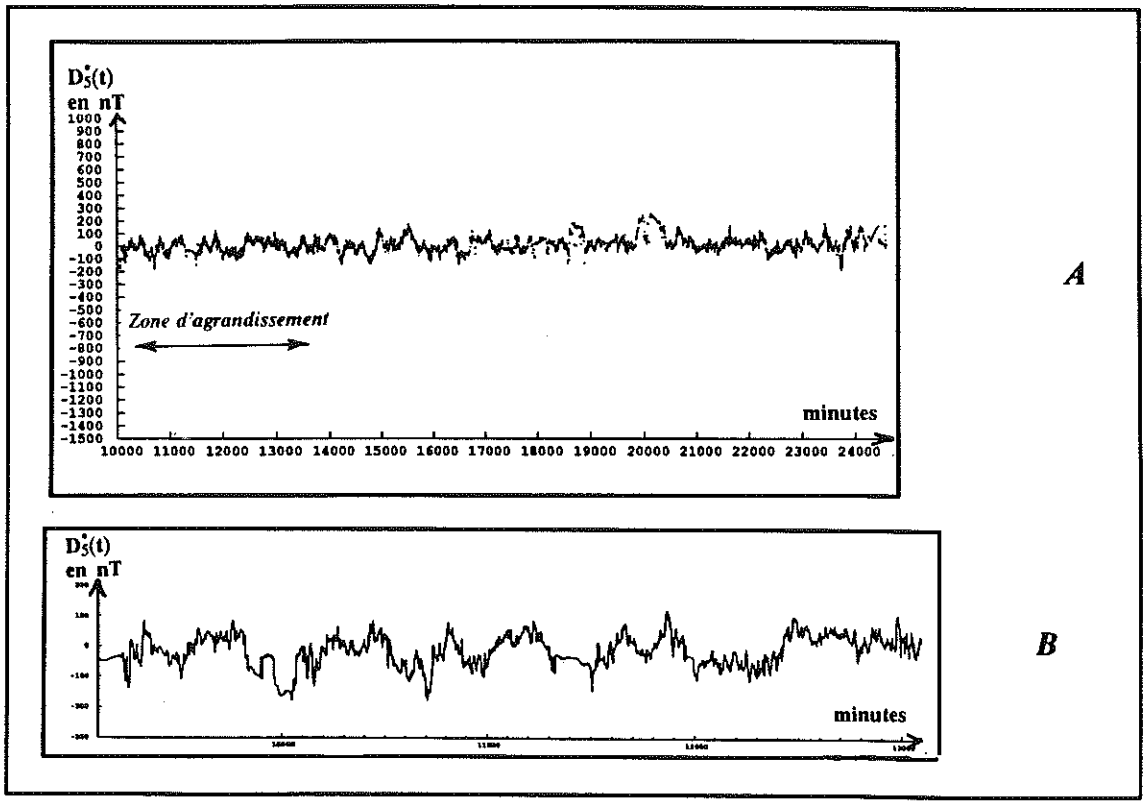


Figure 4.45 : Estimateur et réalité des perturbations (cinquième modèle)

La comparaison des différents variogrammes (figure 4.46) montre une importante réduction de variance : la combinaison linéaire constitutive de l'estimateur contient 40 points au lieu de 20 précédemment . Cette nouvelle estimation n'a pas apporté grand chose, sinon qu'à force de régulariser l'estimateur, on finira par obtenir une variabilité nulle .
Comme l'estimateur spatial résultant n'apporte rien de nouveau, nous ne le présentons pas.

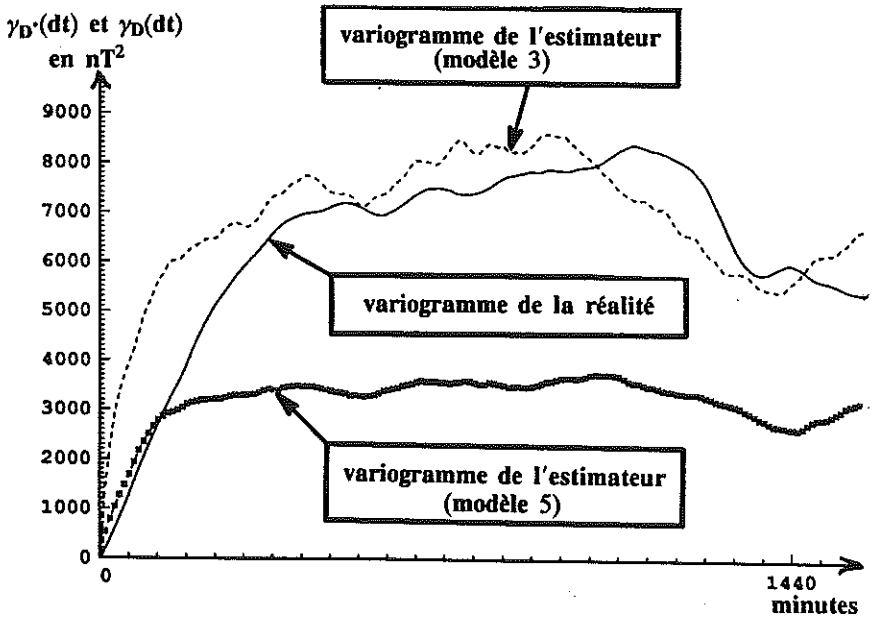


Figure 4.46 : Variogrammes des estimateurs (modèles 3 et 5) et de la réalité

4 .4 .8 Conclusions sur Mohns

La figure 4.47 montre la carte non filtrée et celle obtenue par le troisième modèle . Les avis des spécialistes divergent quant à l'interprétation de ces résultats. Nous pensons qu'aucune de ces 4 estimations n'est acceptable. Si on se remémore le fil de l'analyse, de la modélisation et de l'estimation, on constate qu'à tous les niveaux, le KT présentait de grandes difficultés de mise en oeuvre . Ceci explique sans doute la difficulté d'interprétation des résultats et notre opinion sur le sujet .

Concernant ce dernier cas d'étude, la conclusion essentielle porte sur la géométrie du voisinage et plus particulièrement la configuration $\mathcal{V}_3^x$. En effet, il peut s'avérer dangereux d'appliquer cette géométrie sur des données non périodiques dans le temps car elle régularise l'estimateur autant que l'on veut.
Par contre, s'il y a période temporelle, la réalité des données est un frein à cette régularisation puisque l'analyseur est alors capable de caler les sinusoïdes du système sur les tendances réelles.

Cette constatation expérimentale réduit le KT spatio-temporel à la seule estimation des énergies véritablement "émergentes" du spectre, et ne permet en aucune façon d'estimer un spectre complet. On verra dans le chapitre 5 comment procéder pour estimer des énergies non "émergentes".

Comme les perturbations contiennent d'autres fréquences, l'amélioration la plus immédiate à apporter serait d'utiliser une CG spatio-temporelle dont la contribution dans le temps permettrait de modéliser et de capter les hautes fréquences qui manquent actuellement.

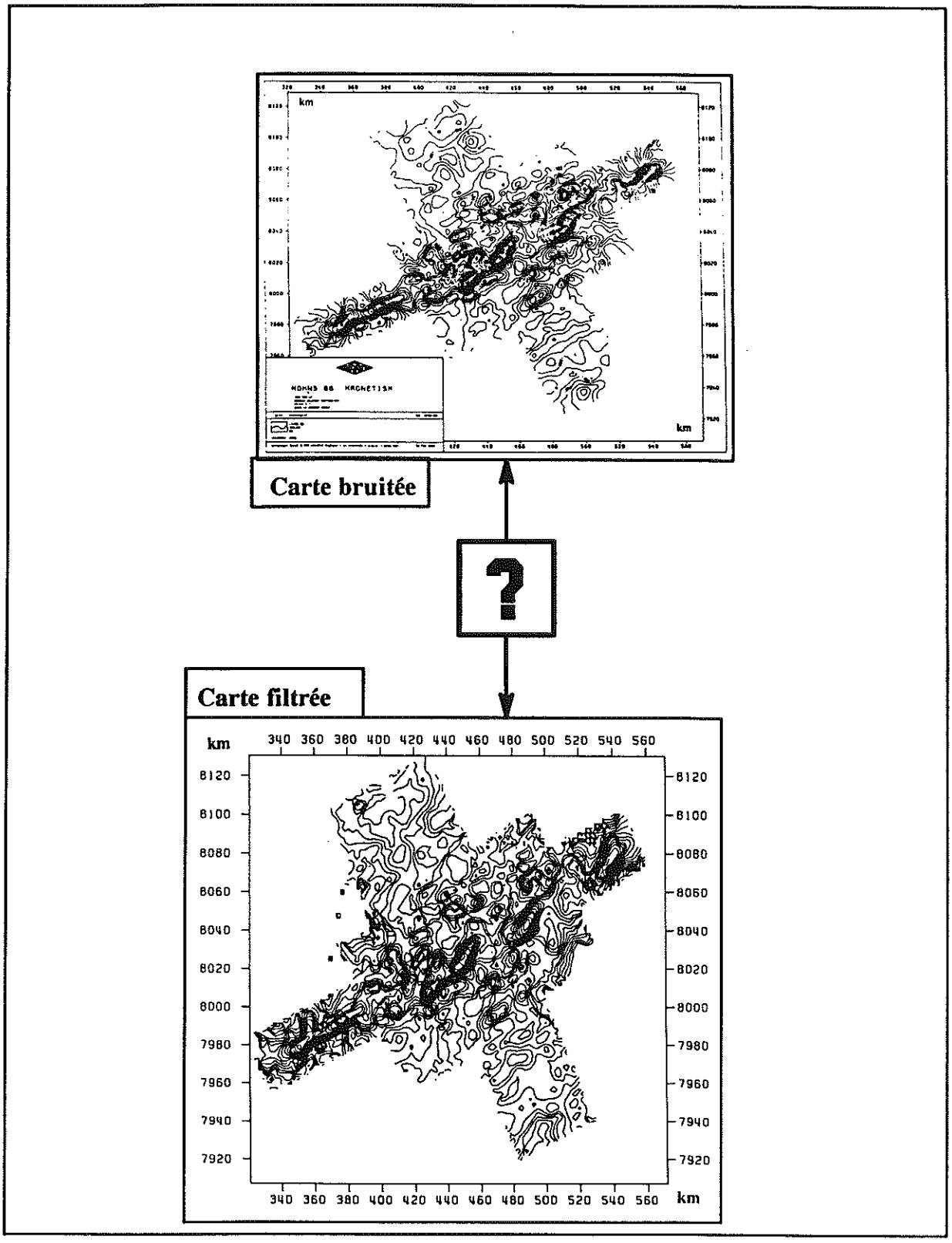


Figure 4.47 : S'agit–il d'un filtrage ou d'un bruitage ?

4 .5 Conclusions du chapitre 4

Le bilan global est positif, puisque lorsque, comme pour Seaperc et Krakatao, les périodes temporelles sont marquées, il est possible de les atteindre par le Krigeage Trigonométrique . Les faits marquants, dans notre manière d'appréhender ces oscillations sont :

- La géométrie de type V_3^x est la plus adaptée au problème, malgré son caractère "sautant" .
- S'il n'y a pas de périodicité manifeste dans le temps, cette géométrie peut être dangereuse dans la mesure où elle régularise excessivement l'estimateur .
- Les principaux problèmes rencontrés lors de la mise en oeuvre du KT sont des problèmes d'instabilités numériques liées à un échantillonnage spatio-temporel imparfaitement dominé .
- Malgré ces instabilités, il a été possible de correctement séparer l'énergie temporelle de l'énergie spatiale. Ceci s'est concrétisé par des variances expérimentales (pour Seaperc) du même ordre de grandeur que celles rencontrées lorsque le signal d'entrée du système est la perturbation elle même .

- Nous avons une fois de plus été confronté à des problèmes de recouvrement spectral. Ce fait expérimental illustre l'importance des fluctuations de l'estimation qui apparaissent lorsque l'analyse et l'estimation sont conduites en voisinage glissant . On peut dire en quelque sorte que la pratique glissante correspond à une dépense d'énergie qui se retrouve dans l'estimateur lorsque les caractéristiques du voisinage sont inadaptées au phénomène.
  Les propriétés de cette énergie parasitaire sont:
  - de donner du détail à l'estimateur ,
  - de déformer ce détail, surtout en l'amplifiant .

  Le fait marquant est que cette énergie apparaît d'une manière naturelle dans l'estimateur et qu'il ne lui correspond aucune modélisation . La conséquence est qu'on ne la domine que par le biais des caractéristiques de voisinage.

CHAPITRE 5

AUTRES APPLICATIONS (IMAGERIE)

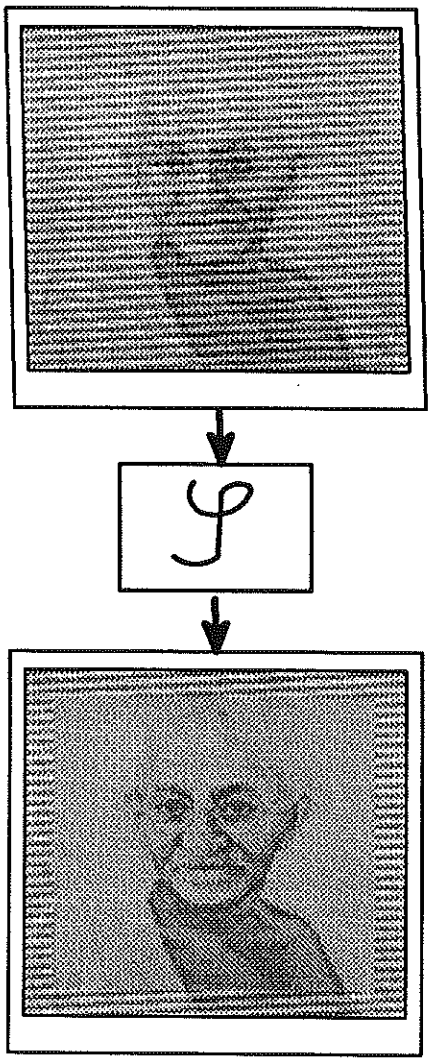
RÉSUMÉ

Ce chapitre offre l'occasion de présenter deux autres applications du Krigeage lorsqu'il est combiné à des fonctions trigonométriques. Elles concernent :

- Le filtrage d'une image à balayage.
  Ce problème est homologue de celui des perturbations diurnes en magnétisme, mais se présente dans des conditions expérimentales beaucoup plus favorables.
  Bien que cette expérience soit anecdotique, elle confirmera le fait que les principaux problèmes rencontrés lors du filtrage du magnétisme, étaient dus à l'échantillonnage des données sur profil.
- L'estimation des hautes fréquences spatiales d'une image.
  C'est à l'aide d'un système d'Analyse Krigeante Trigonométrique que nous estimerons les zones de transition rapide d'une image.
  Outre le fait que cet exemple représente la potentialité d'applications nouvelles, il permet de montrer l'intérêt d'une spécification fréquentielle en covariance (et non en dérive comme pour le Krigeage Trigonométrique), ainsi qu'une indication de la manière dont on peut kriger des énergies qui n'émergent pas du spectre.

5.1 Filtrage d'une image à balayage page 151

5.2 Estimation de hautes fréquences spatiales page 159



Chapitre 5

AUTRES APPLICATIONS (IMAGERIE)

5 .1 Filtrage d'une image à balayage

Dans le but d'illustrer le fait que les principales difficultés de mise en oeuvre du KT sont essentiellement liées à l'anisotropie de l'échantillonnage, et au caractère non stationnaire du modèle, nous présentons dans ce qui suit une application de ces techniques à une image à balayage.

L'image de base de ces expériences est l'image "Gandhi" (figure 5.1).

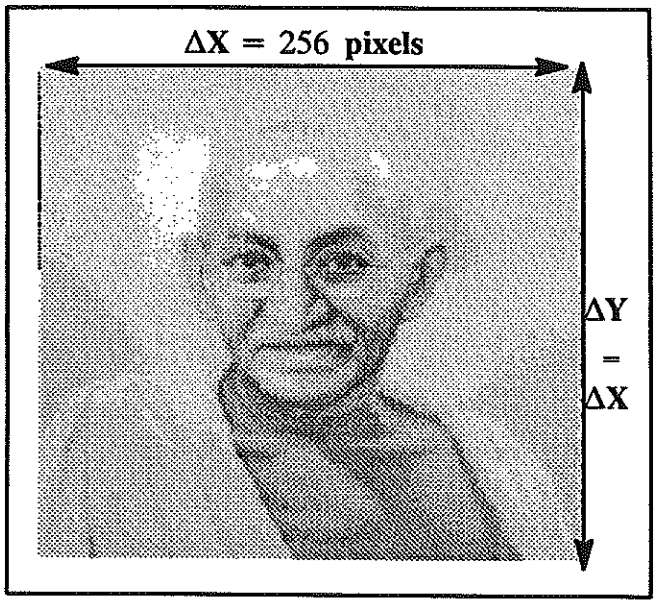


Figure 5.1 : Image initiale

Cette image est représentée en teinte de gris variant de $\emptyset$ à 255 ^{u}, et elle est constituée de 256x256 pixels . Elle nous a aimablement été fournie par le Centre de Morphologie Mathématique de Fontainebleau (ref **[I-4]**) et nous l'en remercions .

La figure 5.2 représente le variogramme de ce visage, pour les deux directions OX et OY . Ces variogrammes ont fait l'objet d'un ajustement isotrope et sur les 30 premiers pixels (figure 5.3).
Le modèle inféré est la somme de 3 structures stationnaires :

$\gamma_{image}(h) = \sigma_{HF}(h) + \sigma_{MF}(h) + \sigma_{BF}(h)$

avec :

$\sigma_{HF}(.)$: modèle cubique de portée 8 ^{pix} et de palier 60.23

$\sigma_{HF}(.)$: modèle cubique de portée 20 ^{pix} et de palier 134.8

$\sigma_{HF}(.)$: modèle sphérique de portée 50 ^{pix} et de palier 346.7

Si les index "HF", "MF" et "BF" sont mentionnés, c'est en raison de l'interprétation qui est faite de ces structures :

HF : (Hautes Fréquences spatiales), représente les zones de transition rapide de l'image.

MF : (Moyennes Fréquences spatiales), représente les structures plus étendues de l'image.

BF : (Basses Fréquences spatiales), représente le fond de de l'image .

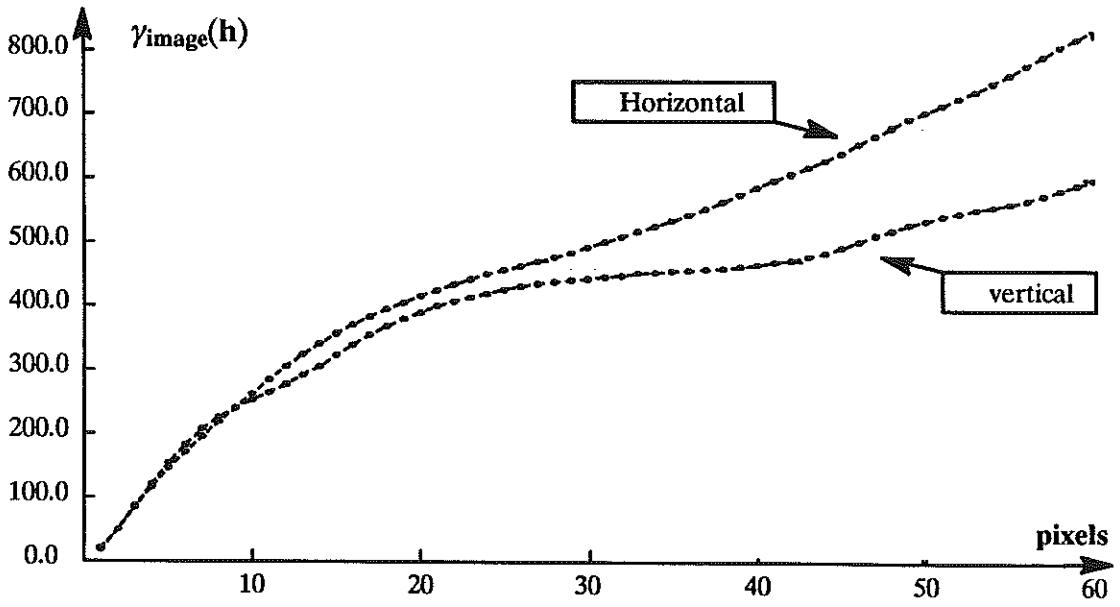


Figure 5.2 : Variogrammes de l'image "Gandhi"

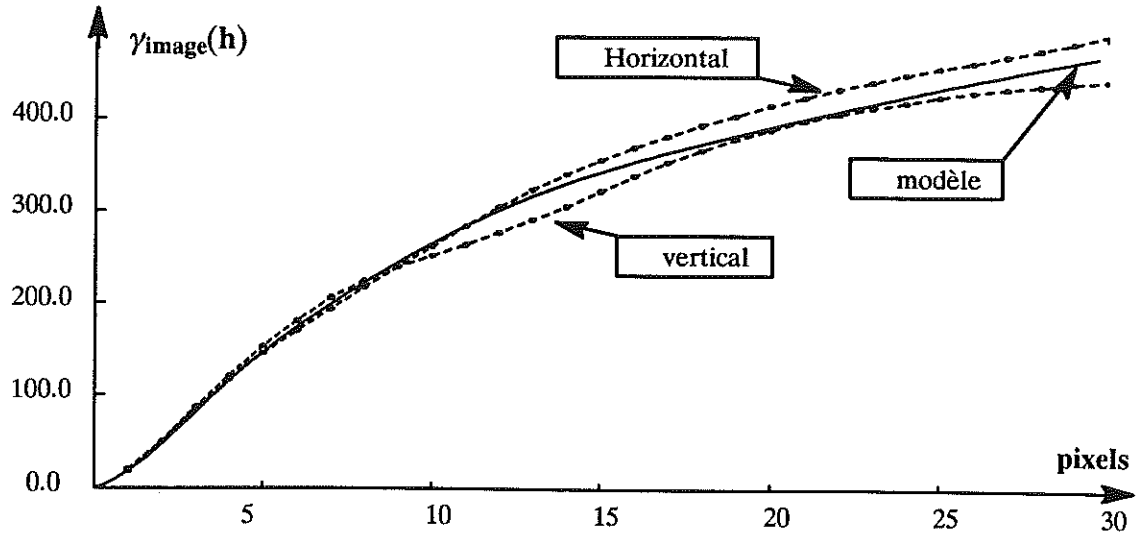


Figure 5.2 : Ajustement des variogrammes de l'image

L'objet de la section 5.1 est de présenter l'analogue du problème des perturbations diurnes, mais dans le cas d'une image à balayage , et dans des conditions expérimentales beaucoup plus favorables . Le travail qui suit peut donc s'interpréter de deux manières :

- Ce qu'on gagne en facilité de filtrage par rapport au magnétisme, lorsque l'image est (considérée comme) stationnaire et la perturbation strictement périodique .
- Un exemple d'application du KT à des images à balayage, et de ce fait le potentiel de cette méthode .

5 .1 .1 Présentation du problème

Les figures 5.4 et 5.5 représentent ce qu'il advient de l'image initiale lorsqu'elle est perturbée par une sinusoïde de période 120 (respt. 512).

La régionalisation de la perturbation est le temps de balayage horizontal exprimé en pixels. On considère que le temps de retour du spot vidéo est nul.

Dans le premier cas (T = 120 pixels), il existe un déphasage entre la période de la perturbation et le temps de balayage d'une ligne de l'image, de sorte que les linéations perturbatrices sont obliques.

Elles deviennent horizontales (figure 5.5) lorsque la période de la perturbation est un multiple du temps de balayage d'une ligne.

On reconnaîtra dans cette maquette un phénomène bien connu en télévision qui apparaît lorsque l'onde hertzienne porteuse de l'information est perturbée par des parasites cycliques (interférences dues au courant d'alimentation du récepteur ou à des véhicules mal anti-parasités, etc...)

5 .1 .2 Liens avec le problème des perturbations en magnétisme

Bien que comparable au problème des perturbations diurnes, on se trouve ici dans une circonstance favorable où :

- A l'échelle de 30 pixels, l'image spatiale initiale peut être considérée comme stationnaire et le recours à une modélisation par FAI-k n'est pas utile.
- La perturbation dans le temps de balayage est strictement périodique et non seulement à "tendance" périodique comme c'est le cas pour le magnétisme.
- Une translation dans le plan en est une également dans le temps de balayage .
  En effet, si on interprète chaque ligne de balayage comme un "profil", on constate que ceux-ci vont tous de la gauche vers la droite , ce qui n'était pas le cas des données de la géophysique.
  Il en résulte qu'en translatant la fenêtre d'estimation dans l'espace, on la translate également dans le temps.
  Comme de plus nous avons affaire à un échantillonnage périodique strictement, nous n'aurons à inverser le système linéaire qu'une seule fois pour obtenir sa solution que l'on convoluera alors sur chacun des pixels qu'on souhaite filtrer.
  Le gain de temps est considérable.

Fort de ces remarques, il est intéressant de voir ce que donne le KT lorsque les conditions expérimentales sont aussi favorables . Comme le bruit est strictement périodique, il faut bien reconnaître que ces expérimentations sont quelque peu anecdotiques.

Une petite circonstance atténuante : la phase et l'amplitude de la sinusoïde sont inconnues...

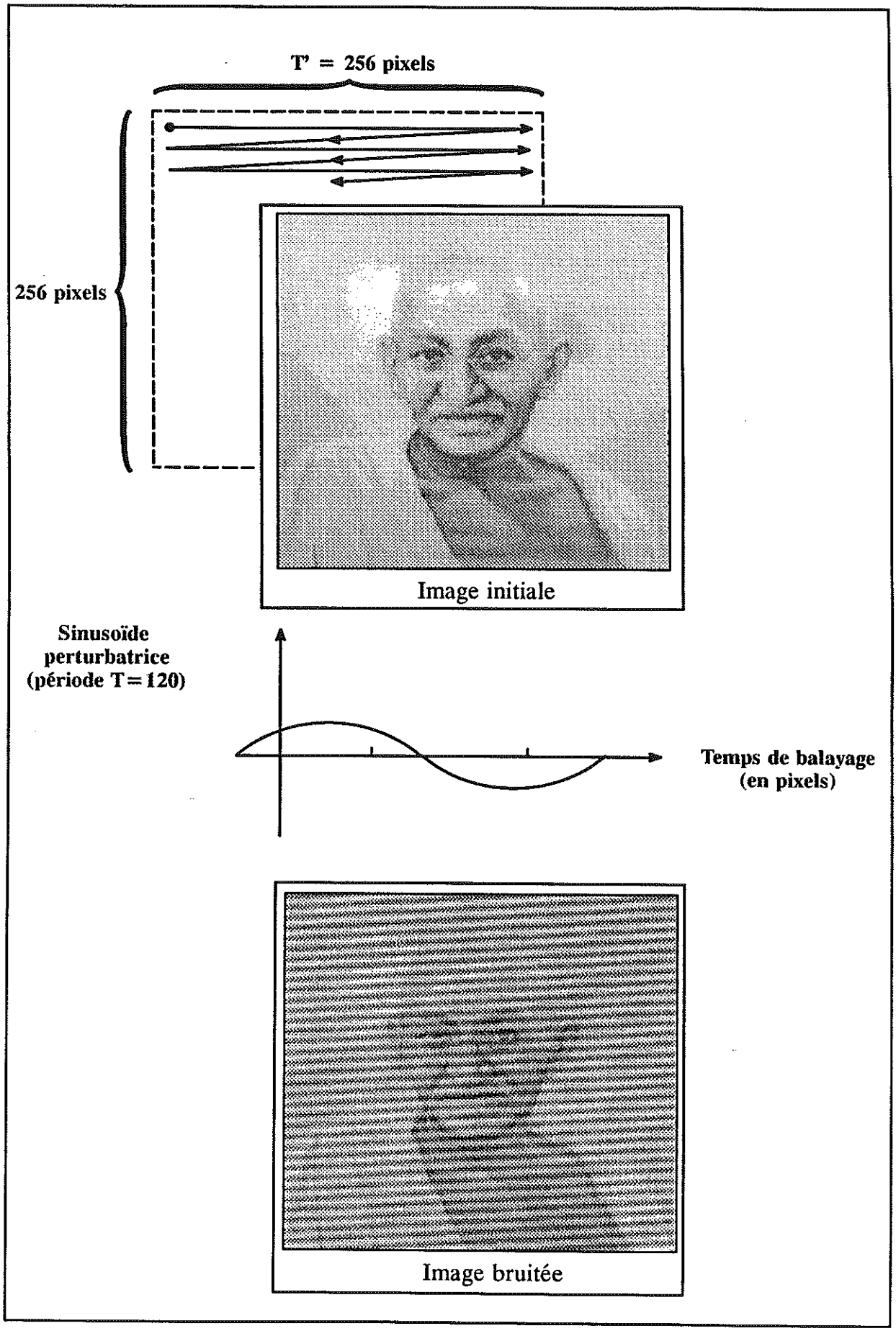


Figure 5.4 : Simulation d'une perturbation vidéophonique (période perturbatrice : 120 pixels)

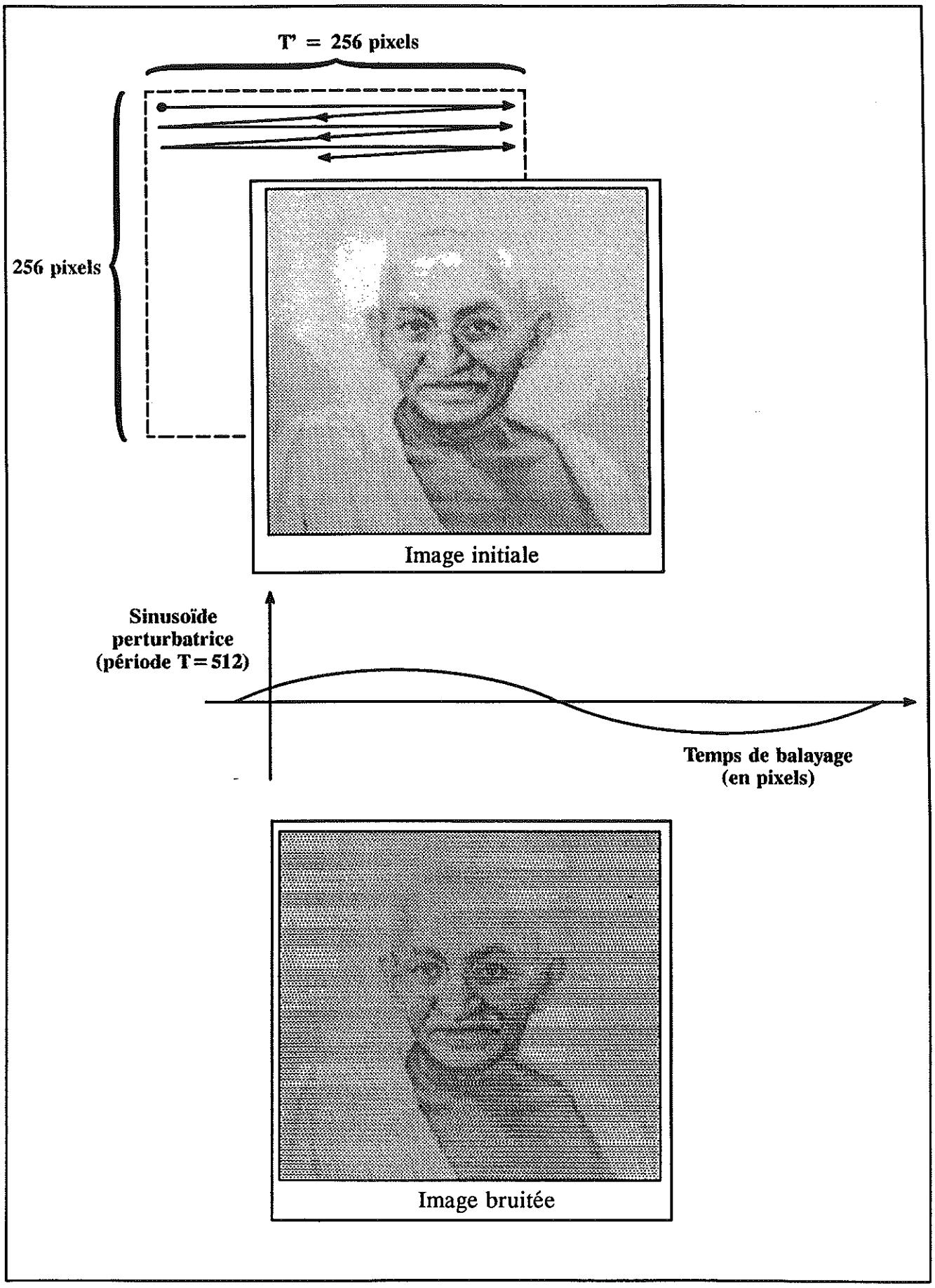
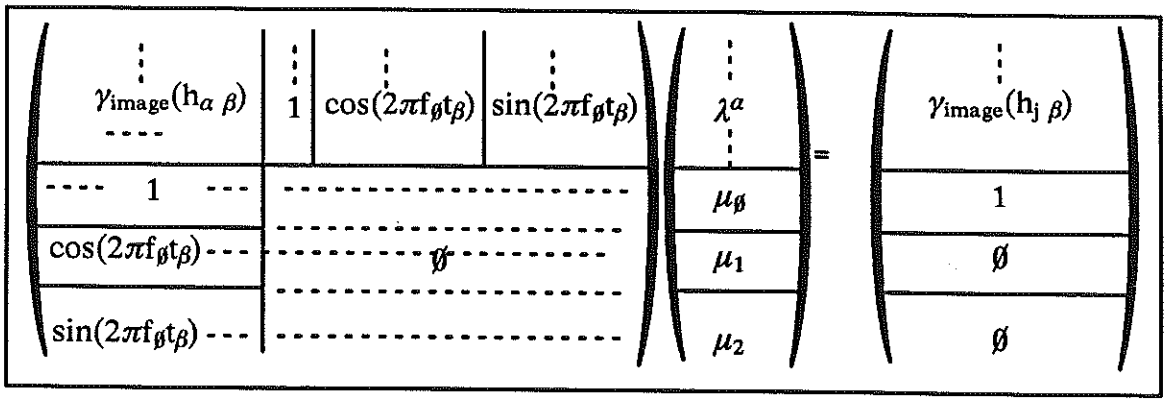


Figure 5.5 : Simulation d'une perturbation vidéophonique (période perturbatrice : 512 pixels)

5 .1 .3 Système linéaire utilisé

Le système de filtrage utilisé est un désormais classique système de KT, dans un contexte de FA spatiale à variogramme .



5 .1 .4 Résultats du filtrage de la période T = 120 pix

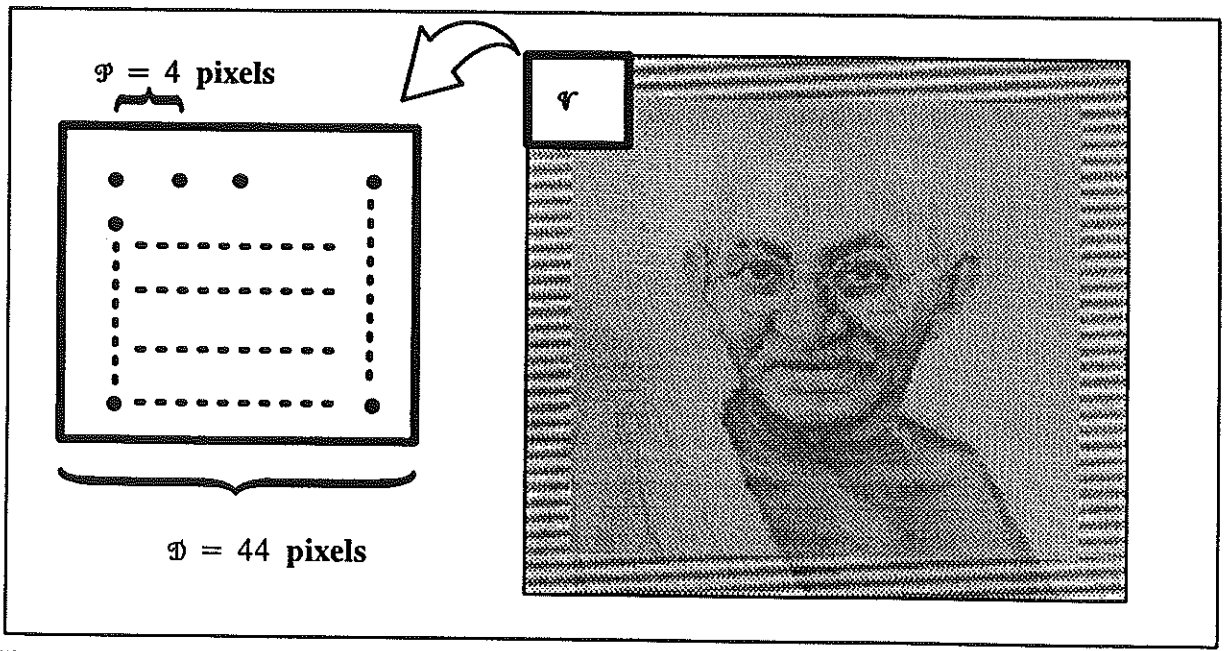


Figure 5.6 : Restauration obtenue par krigeage trigonométrique

La figure 5.6 représente le résultat de l'estimation lorsque celle-ci est conduite avec un voisinage glissant dont les caractéristiques sont :

- $\mathcal{V}$: maille régulière spatiale centrée sur le pixel ciblé par l'estimation .
- Le pixel ciblé par l'estimation appartient à $\mathcal{V}$.
- $\mathcal{P}$ = 4 pix .
- Grille de 11 par 11 points.

Sur cette figure, l'existence d'une bordure non filtrée, (outre le fait qu'elle mette opportunément le filtrage en évidence...), est due aux sections horizontales et verticales du diamètre (44 pix) qui sont impliqués par ce voisinage (on s'est imposé d'estimer tous les pixels avec la même géométrie pour $\mathcal{V}$).

Le filtrage à nécessité 1/10 ème de seconde CPU sur Sun .

La figure 5.7 représente les évolutions comparées de la sinusoïde perturbatrice et de son estimation .

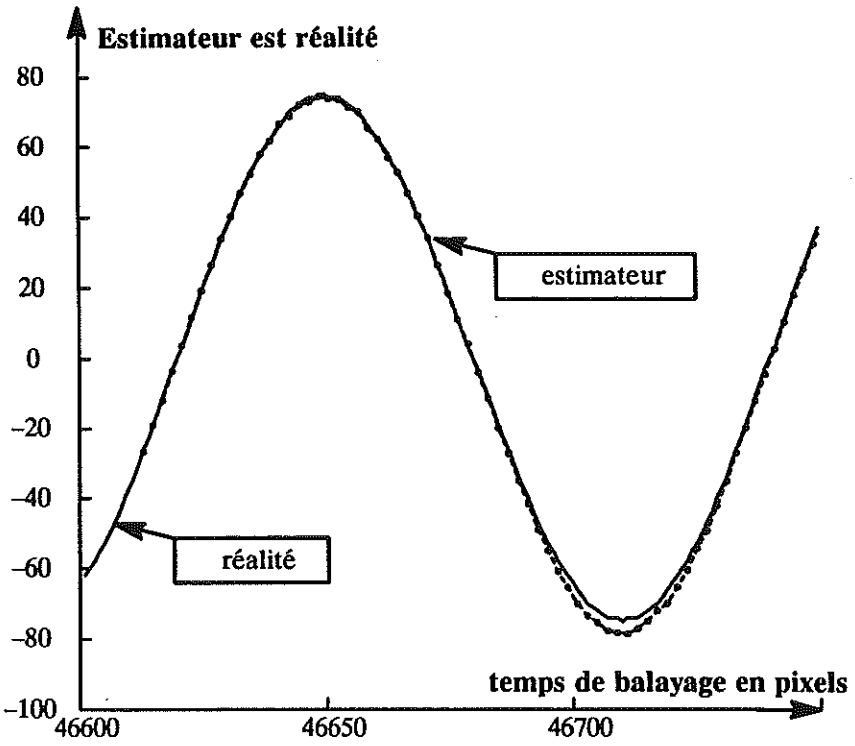


Figure 5.7 : Estimateur et réalité

La qualité des résultats se passe de commentaire .

5 .1 .5 Résultats du filtrage de la période T= 2x256 pix

La figure 5.8 représente le résultat de l'estimation lorsque celle-ci est conduite avec un voisinage glissant dont les caractéristiques sont :

- $\mathcal{V}$: maille régulière spatiale centrée sur le pixel ciblé par l'estimation .
- Le pixel ciblé par l'estimation appartient à $\mathcal{V}$.
- $\mathcal{P}$ = 1 pix .
- 3 points en X , 3 points en Y.

La figure 5.9 représente les évolutions comparées de la sinusoïde perturbatrice et de son estimation .

La qualité des résultats se passe de commentaire .

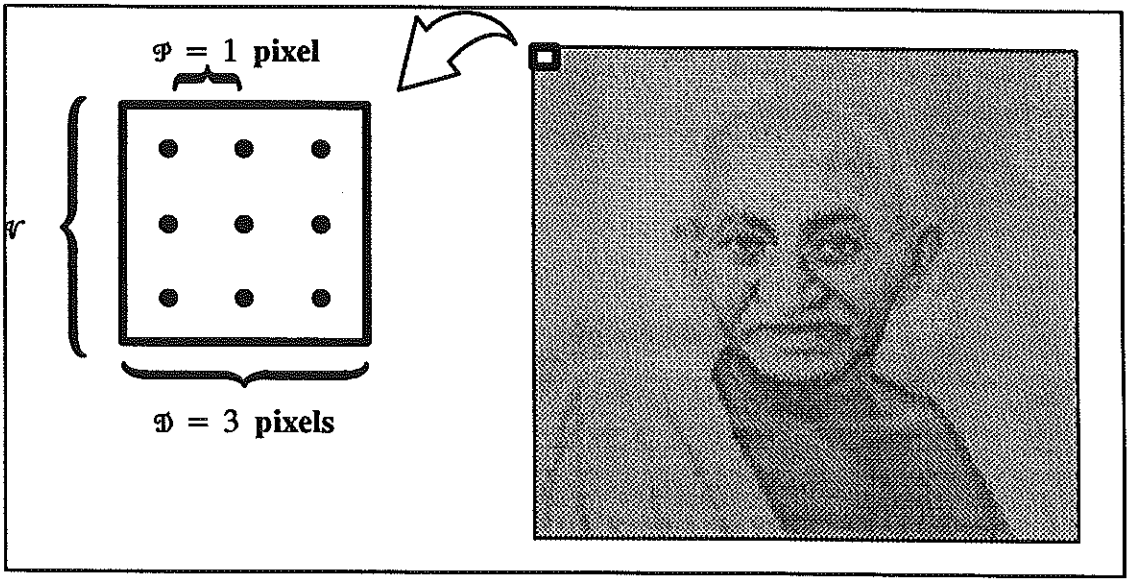


Figure 5.8 : Restauration obtenue par krigeage trigonométrique

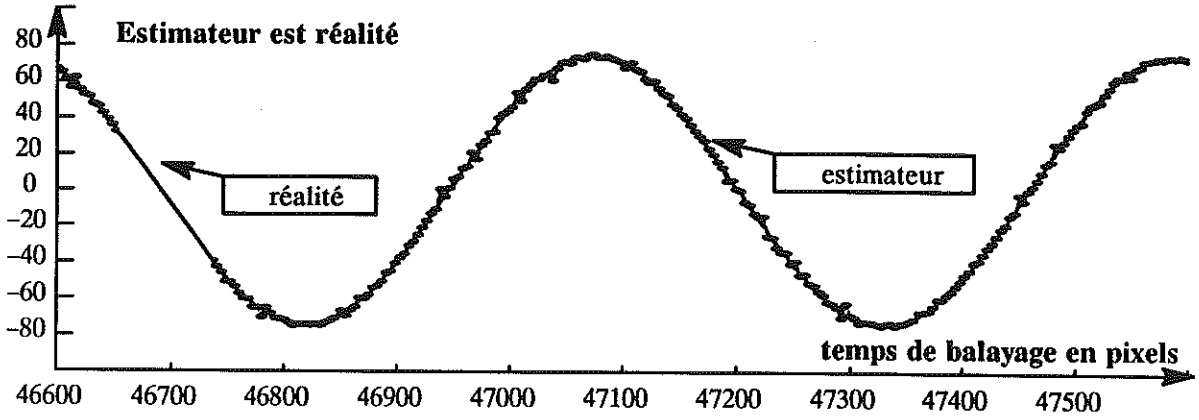


Figure 5.9 : Estimateur et réalité

5 .1 .6 Conclusions sur l'image à balayage

Les résultats obtenus par KT sont très satisfaisants . Ils montrent que lorsque nous n'avons aucun problème de stabilité numérique, cette méthode est très puissante .

Mais il faut bien reconnaître que la stricte périodicité du bruit relègue ces expériences au rang d'anecdotes. Pour les applications pratiques, il conviendrait de renouveler ce travail sur un signal perturbé par une dizaine de composantes. et de voir si l'estimation s'avère toujours aussi performante. De plus, nous n'avons conduit aucune analyse puisque nous avons utilisé directement le variogramme (inaccessible en général) de l'image non bruitée. Il conviendrait de voir si les fréquences temporelles et le variogramme initial demeurent plus accessible qu'une CG en FAI-k.

Quoi qu'il en soit, ces expériences de filtrage illustrent bien où se situent la majorité des problèmes rencontrés (dans l'échantillonnage) ; et qui sait sont peut être le point de départ d'autres applications du KT.

5 .2 Estimation de hautes fréquences spatiales d'une image

En ce qui concerne le magnétisme spatio-temporel, un choix arbitraire a été adopté tout au fil de ce travail : celui, dans le modèle, de mettre l'information fréquentielle (les sinusoïdes temporelles) en termes de dérive. Il en résultait l'opérateur KT de filtrage. Mais nous aurions pu tout aussi bien mettre l'information fréquentielle en covariance et utiliser une covariance spatio-temporelle incorporant (pour la partie temporelle) des fonctions trigonométriques.
On peut se demander alors où se situe la différence, puisque l'une ou l'autre approche est a priori possible. C'est dans ce but qu'un certain nombre d'expériences ne figurant pas dans cette thèse furent conduites. Elles étaient mono-dimensionnelles, et s'appliquaient à compléter par des estimations, une information d'entrée parcellaire (des échantillons de perturbations diurnes). Les conclusions résultantes furent alors que pour un certain diamètre de krigeage, il y avait une équivalence entre les deux approches, et que dans tous les cas, il était préférable de mettre l'information fréquentielle en dérive.

Le cadre très restreint de ces expériences ne permet pas de généraliser ces résultats et d'apporter une réponse formelle à ce problème.

Nous nous proposons dans ce qui suit de montrer à travers un exemple, qu'il peut s'avérer utile de parfois mettre en covariance la spécification fréquentielle. Mais cet exemple possède une autre latitude : celle que l'image n'est pas un PTP puisque l'estimation va être conduite sur l'image Gandhi déjà présentée. Sur celle-ci, nous allons tenter d'atteindre l'objectif suivant : estimer les contours du visage par Analyse Krigeante Trigonométrique (AKT).

Si on se remémore les déboires du cas d'étude Mohns, on comprendra qu'estimer des fréquences (qu'elles soient temporelles ou non) dans un signal qui ne les particularise pas, est un échec si le KT est utilisé dans ce but. Il en résulte que résoudre ici ce problème des hautes fréquences spatiales n'est envisageable que par AKT.

Afin de rendre les résultats plus démonstratifs encore, nous les comparerons à ceux d'une Analyse Krigeante classique.

5 .2 .1 Modèle disponible

On rappelle le modèle de variogramme qui sera utilisé pour mener à bien l'estimation :

$$\gamma_{image}(h) = \sigma_{HF}(h) + \sigma_{MF}(h) + \sigma_{BF}(h)$$

Dans ce modèle, nous nous intéresserons plus particulièrement à la structure $\sigma_{HF}(.)$ dont on rappelle qu'elle est de portée $a = 8$ pix. Cette structure est interprétée comme la variabilité des zones de transition du visage (ie de ses contours).

5 .2 .2 Systèmes linéaires utilisés

Deux systèmes linéaires seront utilisés pour mener à bien l'estimation :

●●Un premier système d'Analyse Krigeante Trigonométrique dans lequel la fréquence du second membre est celle qu'on souhaite capter par l'estimation : $\frac{1}{a}$

Pour plus de détail sur l'analyse Krigeante, on consultera l'annexe B.
Dans le système, l'énergie ξ_a est arbitraire et sa valeur sert en fait de paramètre d'amplification de contraste entre fond et contour. Nous présentons sa forme matricielle :

$$\left(\begin{array}{c|c} \vdots & \vdots \\ \gamma_{image}(h_{\alpha\beta}) & 1 \\ \vdots & \vdots \\ \hline \cdots\ 1\ \cdots & \emptyset \end{array}\right)\left(\begin{array}{c} \vdots \\ \lambda^{\alpha} \\ \vdots \\ \hline \mu_{\emptyset} \end{array}\right) = \left(\begin{array}{c} \vdots \\ \xi_a[1-\cos(\frac{2\pi}{a}h_{j\beta})] \\ \hline \emptyset \end{array}\right)$$

On peut justifier ce système d'une manière heuristique en disant que l'ajustement du variogramme de l'image possède une TFD n'engendrant que des énergies négatives.
Il en résulte que ce variogramme (qui est un modèle) peut être ajusté par une somme de modèles périodiques, dont celui figurant au second membre.
Présenté ainsi, ce système n'est qu'un système d'Analyse Krigeante classique, dans lequel seules des fonctions trigonométriques interviennent

●●Un deuxième système d'Analyse Krigeante dans lequel le second membre possède la structure $\sigma_{HF}(.)$ qui on le rappelle possède une portée $a = 8$ pix. Nous présentons sa forme matricielle :

$$\begin{pmatrix} \gamma_{image}(h_\alpha\ \beta) = & \vdots \\ + \quad \sigma_{HF}(h_\alpha\ \beta) & \\ + \quad \sigma_{MF}(h_\alpha\ \beta) & 1 \\ + \quad \sigma_{BF}(h_\alpha\ \beta) & \\ \hline \text{----} \quad 1 \quad \text{---.} & \emptyset \end{pmatrix} \begin{pmatrix} \vdots \\ \lambda^\alpha \\ \vdots \\ \hline \mu_\emptyset \end{pmatrix} = \begin{pmatrix} \vdots \\ \sigma_{HF}(h_j\ \beta) \\ \hline \emptyset \end{pmatrix}$$

On remarquera que pour chacun de ces systèmes, la moyenne mobile (implicitement krigée) est exclue de l'estimateur .

5 .2 .3 Estimation par Analyse Krigeante Trigonométrique (AKT)

La figure 5.10 montre le résultat de l'estimation lorsque le voisinage possède les caractéristiques suivantes :

- $\mathcal{V}$ est (dans la mesure du possible) circulaire .
- Il contient tous les points contenus dans le cercle .
- Le diamètre est de $\mathfrak{D} = 8$ pix .

Le résultat est assez remarquable .

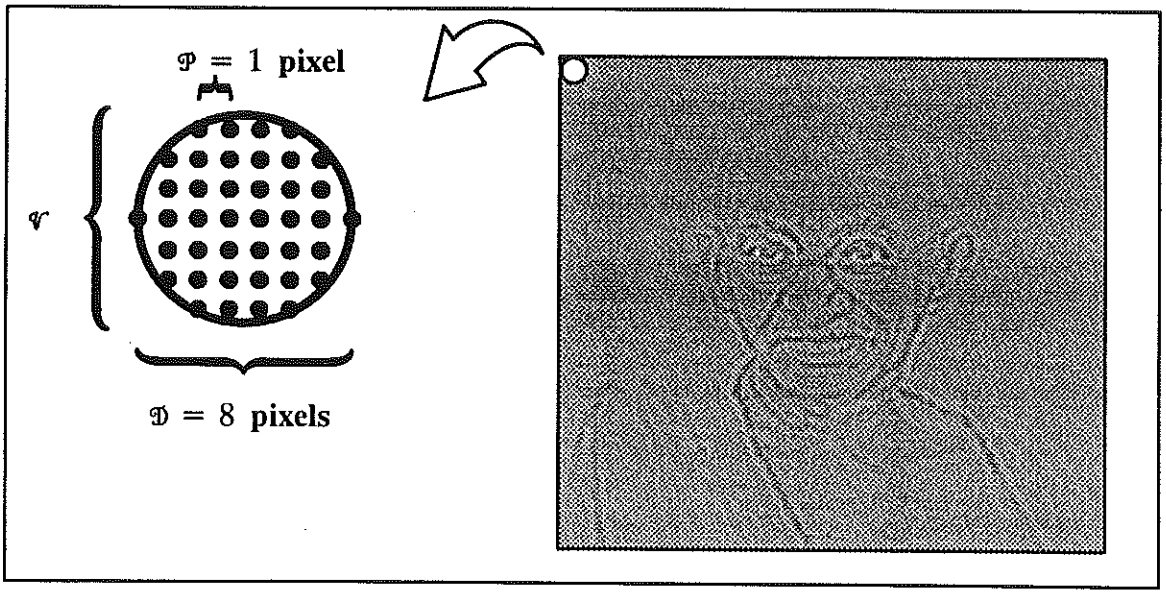


Figure 5.10 : Contours estimés par analyse krigeante trigonométrique (période ciblée : 8 pixels)

5 .2 .4 Estimation par Analyse Krigeante classique (AK)

La figure 5.11 montre le résultat de l'estimation lorsque le voisinage possède les caractéristiques suivantes :

- $\mathcal{V}$: maille régulière spatiale centrée sur le pixel ciblé par l'estimation .
- Le pixel ciblé par l'estimation appartient à $\mathcal{V}$.
- $\mathcal{P}$ = 1 pix .
- 17 points en X , 17 points en Y.

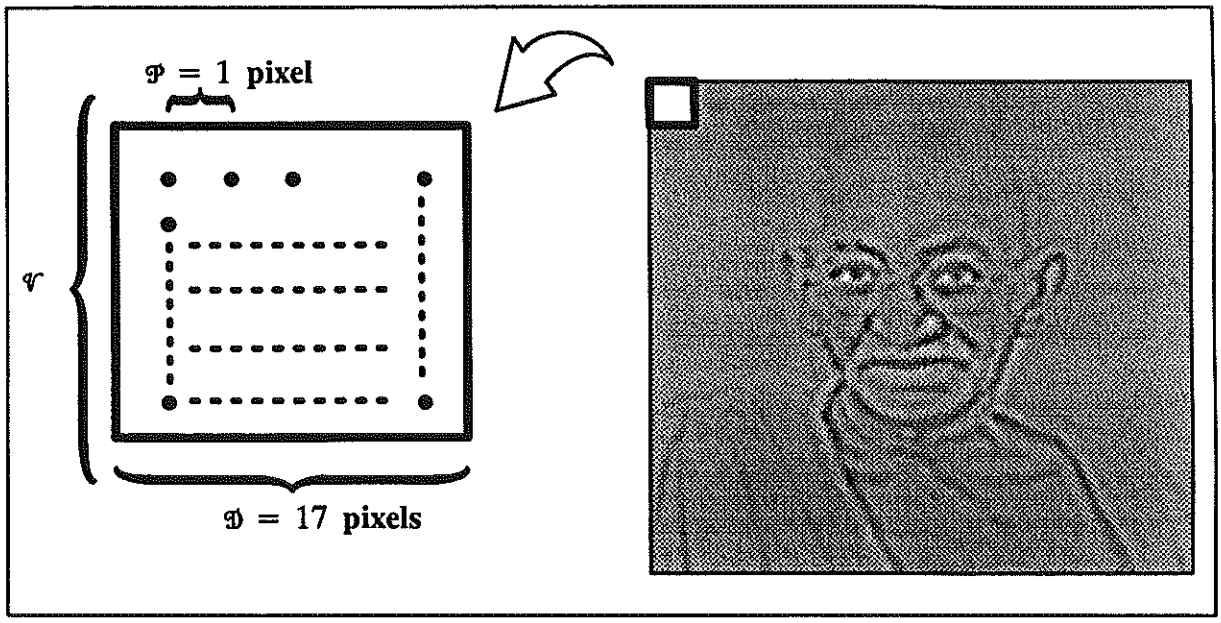


Figure 5.11 : Contours estimés par analyse krigeante classique (périodes ciblées : de 0 à 8 pixels)

On note qu'ici encore, les contours du visage sont parfaitement extraits de l'image initiale .

Comparativement au cas précédent, on constate que les traits sont plus larges, ce qui se comprend aisément puisque le système d'AK donne en théorie à l'estimateur, toutes les périodes comprises entre $\emptyset$ et 8 pix, alors que dans le cas de l'AKT (et puisque le diamètre de Krigeage est résonnant avec la période du système et que la moyenne glissante n'est pas incluse dans l'estimateur) c'est très exactement une TFD 2D glissante qui vient d'être pratiquée (ou plus exactement sa généralisation à un bruit coloré) .

5 .2 .5 Nécessité d'un diamètre adapté

La figure 5.12 présente le résultat obtenu par AKT lorsque le voisinage possède les caractéristiques suivantes :

- $\mathcal{V}$ est (dans la mesure du possible) circulaire .
- Il est constitué de tous les points contenus dans le cercle .
- Le diamètre est de $\mathcal{D}$ = 16 pix, soit le double de la période ciblée.

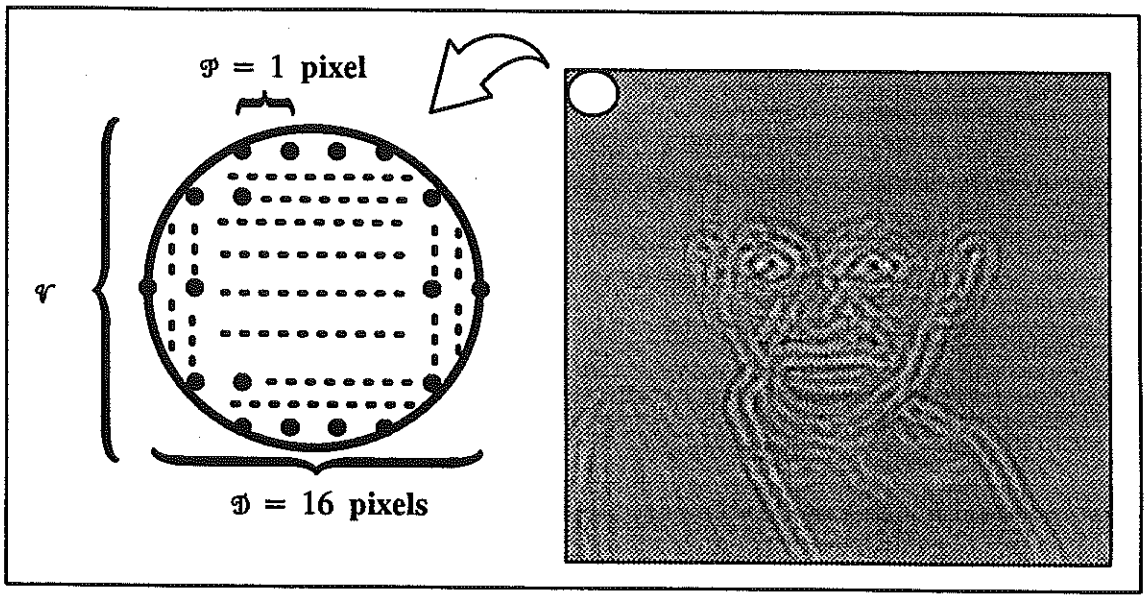


Figure 5.12 : Création d'un moiré par voisinage inadapté (période ciblée : 8 pixels, diamètre : 16 pixels)

On note l'apparition d'un moiré .

Ce fait expérimental s'explique comme suit : lorsque la fenêtre d'estimation est située à 13 ou 14 pixels d'une zone de transition, elle capte déjà l'énergie de cette transition , cale sa sinusoïde sur cette énergie et l'affecte au pixel ciblé . Il en résulte que les contours du visage se trouvent très exactement dupliqués par dilatation, ce qui donne le résultat que l'on sait .

On peut imaginer diverses applications à cette curiosité expérimentale :

- Puisqu'on sait créer un moiré, on sait filtrer un moiré "naturel". D'où la potentialité de filtrage de moiré dans une image.
- En répétant une telle AKT plusieurs fois, puis en filtrant tous les contours inclus dans le contour extrémal, on peut obtenir une dilatation de la figure initiale .

5 .2 .6 Conclusions sur l'estimation des contours

Ces expériences sont moins anecdotiques que celles des images à balayage, mais nous confrontent à un autre problème : on ne sait plus trop exactement ce qui est estimé lorsque de tels systèmes sont manipulés. Si le résultat parle de lui-même (les contours du visage apparaissent nettement), son obtention procède de deux choix arbitraires :

- Ne pas incorporer la moyenne mobile dans l'estimateur,
- Choix arbitraire de la fréquence qui est mise dans le second membre.

Pour que ces expériences d'AKT trouvent une utilisation en pratique, il conviendrait donc de les réitérer pour diverses fréquences (spatiales) et d'analyser la sensibilité des résultats à ce dernier paramètre.

CONCLUSIONS

Sur le filtrage des perturbations diurnes en magnétisme page 165

Sur la généralisation à d'autres PTP page 170

Sur la complémentarité spatial–spectral page 171

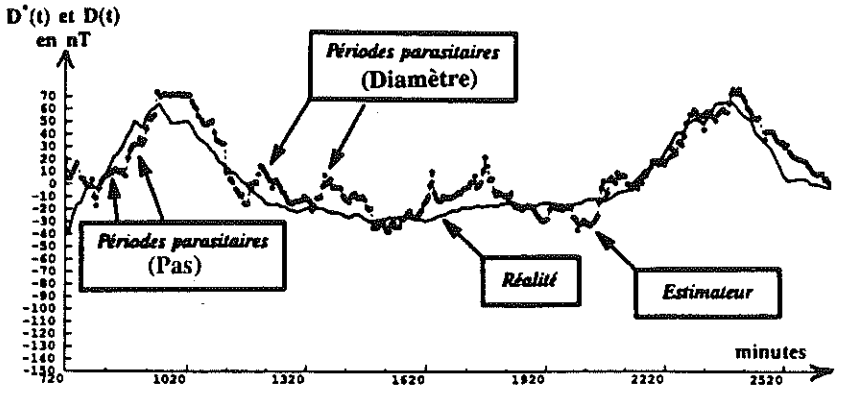
Sur le krigeage de coefficients de Fourier page 172

Sur le caractère esthétique et fascinant du PTP page 173

Conclusions

Sur le filtrage des perturbations diurnes en magnétisme

Le premier fait marquant dans notre approche du magnétisme spatio-temporel est la difficulté à détecter des fréquences temporelles (chapitre 3).
La raison essentielle nous semble être la configuration des échantillons, très proches dans l'espace (et dans le temps), et ceci sur des profils relativement très éloignés les uns des autres. Lors de la constitution du voisinage, cet alignement naturel des points informants rejaillit sur la géométrie des points sélectionnés qui possèdent alors une tendance à respecter l'alignement initial. Or, on l'a vu, les droites de points sont une géométrie incompatible avec l'inférence d'une FAI-k pour k strictement positif. Pour combattre cette tendance à l'alignement, nous avons donc dû élaborer des contraintes lors de la constitution du voisinage, et nous avons été conduit à tester diverses configurations. Il s'est avéré que la plus apte à assurer une isotropie spatiale est celle dénommée ∇_3^x qui finalement est une configuration très proche d'une maille régulière bi-dimensionnelle de pas égal à l'interdistance entre deux profils. Occultons momentanément l'échantillonnage temporel résultant. Le fait que nous ayons été conduit à une telle maille spatiale suscite certaines interrogations. Ainsi, on peut s'interroger sur l'impact de cet échantillonnage sur l'image que nous avons alors du magnétisme, car on sait que dans ces conditions de maille régulière, un pas inadapté au spectre du phénomène nous expose à d'importants problèmes de recouvrement spectral et donc à une déformation de l'image que nous avons de la réalité.
Une autre interrogation concerne la coupure des hautes fréquences spatiales que ce pas de maille implique, c'est à dire la perte du détail de toutes les structures du magnétisme plus petites que ce pas (élevé au carré).
Et enfin, une troisième interrogation persiste : quel est l'impact exact sur l'estimation, de "l'effet sautant" du voisinage. On rappelle que ce phénomène est dû au fait que lorsque le voisinage glisse sur le champ, il ne se renouvelle pas continûment mais par sauts, puisque cette géométrie possède une forte tendance à ne se constituer que des points frontières des pseudo-profils successifs. Je crois bon à ce propos de rappeler une figure déjà rencontrée dans le chapitre 4 : celle d'un estimateur D*(t) des perturbations D(t), sur lequel on peut retrouver les diamètre et pas du voisinage qui ont conduit l'estimation :



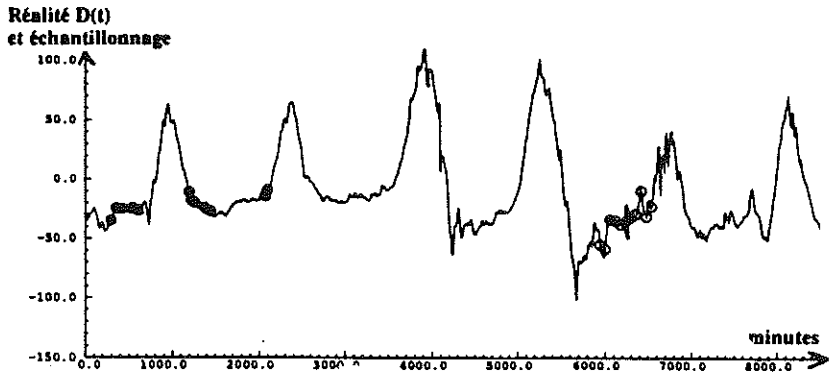
Cette figure illustre parfaitement le fait que supprimer cet effet sautant devrait constituer une amélioration suffisante pour juger que la méthode est aboutie, et c'est à ce dernier problème que nous avons une réponse concrète à apporter : celle de pratiquer un séquencement glissant, ie de faire en sorte qu'effectivement une maille quasi régulière (dans le plan) puisse être adoptée, mais qu'elle glisse avec la cible de l'analyse (ou de l'estimation) de sorte qu'elle n'entraîne pas les fluctuations qu'on observe et aussi qu'elle ne remplace pas les 10000 points informants initiaux par une centaine de points frontières comme c'est le cas actuellement.
Quant à la réponse à apporter aux deux premières interrogations (recouvrement spectral entraînant une déformation des basses fréquences, coupure spectrale entraînant une suppression des hautes fréquences, le tout du fait d'une maille régulière), il semblerait que ce soit dans les techniques de Théorie du Signal qu'il faille chercher quelques éléments de réponse. On pensera en effet à l'élaboration de "filtres anti-repliement" qui sont des pondérations arbitraires affectées aux données du voisinage, et ceci afin de minimiser les effets de recouvrement. Mais le problème tient précisément dans "l'arbitraire" de cette pondération car en théorie, le filtre idéal dépend du spectre de la variable.. Or, nous travaillons en termes de FAI-k, et ce spectre n'existe pas dans cette théorie! On peut imaginer d'associer un spectre à la représentation localement stationnaire qui est utilisée de la FAI-k, mais cela nécessiterait (comme pour les covariances généralisées) de ne manipuler que des combinaisons linéaires autorisées d'ordre k, et de voir quel filtre leur associer ... si tant est que l'on puisse parler de recouvrement spectral pour de tels êtres mathématiques ...
Il semblerait donc que ce problème de recouvrement soit inextricable (ie pas de réponse à la première interrogation) de même qu'il semble inévitable de perdre du détail sur la variable si l'on choisit de travailler sur une maille spatiale avec un pas trop lâche.
Concrètement donc, je ne vois qu'une seule amélioration pour mieux détecter les fréquences temporelles et les estimer : supprimer l'effet sautant du voisinage (je ne rappelle pas ici les quelques suggestions informatiques que l'on trouve dans les chapitres 3 et 4).

Supposons que cette dernière amélioration soit acquise. Même dans ce cas favorable, on se trouvera toujours confronté à ce fameux problème de la "coplanarité" dans l'espace-temps de la route d'échantillonnage du bateau. Cette circonstance nous empêche de dissocier l'échantillonnage spatial de l'échantillonnage temporel, ce qui a pour conséquence immédiate qu'épisodiquement et d'une manière difficilement prévisible, l'échantillonnage temporel est désastreux, avec pour résultats :

- Un système numérique rendu instable par des colonnes trigonométriques linéairement dépendantes. Ceci concerne donc en quelque sorte l'échantillonnage du modèle, et plus exactement de la série de Fourier temporelle associée à celui-ci.

- Un échantillonnage temporel qui ne détectera aucune tendance sur les données, à cause par exemple d'un pas trop lâche en regard des oscillations.

On se remémorera à ce propos les échantillonnages temporels observés pour certaines estimations particulièrement "explosives" et dont je rappelle ci dessous un graphique déjà rencontré :



Sur cette figure, on note que le pas temporel est tellement lâche que D(t) n'est échantillonné que la nuit, zone temporelle où précisément les perturbations fluctuent peu, et de toute façon d'une manière non sinusoïdale. Comment dans ce cas caler une sinusoïde sur de tels échantillons, ce qui est l'opération qu'on tente de faire lors des analyse et estimation ?
Bien souvent d'ailleurs, ces deux phénomènes : matrice non inversible et pas inadapté aux fluctuations temporelles, sont liés et simultanés, ce qui explique peut être la violence des "explosions" de l'estimation. Une résonance dans l'instabilité, en quelque sorte...
Il se pose alors la question du remède à apporter pour éviter ces circonstances critiques qui peuvent survenir épisodiquement ou très régulièrement, selon la configuration de la route du bateau. On peut en réponse imaginer de réduire le pas d'échantillonnage, afin d'assurer un meilleur échantillonnage temporel. Mais c'est précisément ce genre de considérations qui nous ont conduit à la géométrie de type $\mathcal{V}_4^{x\ t}$ qui n'est qu'une superposition d'une maille spatiale de type $\mathcal{V}_3^x$ (lâche dans le temps) et d'une maille temporelle centrée sur la cible (et plus resserrée).
Et c'est là également que la "coplanarité" du problème nous montre les limites de ce que nous pouvons faire : ces deux mailles (spatiales et temporelles) sont incompatibles puisqu'en utilisant $\mathcal{V}_4^{x\ t}$, nous avons assisté à un "duel d'influence" entre cette maille temporelle (qui devient une droite dans l'espace) et cette maille spatiale (qui se réduit à deux ou trois points dans le temps). Ce "duel", que nous avions qualifié de "contraste d'échelle", a lieu dans l'espace, et aboutit à ce que, comparativement à la droite des points contraints temporellement, la maille des points contraints spatialement se réduit à un amas de points. Autrement dit, ce duel est gagné par les contraintes temporelles. Mais alors, on a de nouveau affaire à des instabilités numériques puisqu'une droite de points alignés ... etc ...L'impasse.
Tout cela pour dire que finalement, la seule amélioration envisageable, c'est la suppression de cet effet sautant, et que je crois que ce mode de filtrage, lorsqu'il est appliqué à de tels échantillonnages spatio-temporels, nous confrontera toujours à quelques explosions et instabilités. Il en résulte que cette méthode ne sera jamais "presse-bouton" et nécessitera toujours de regarder avec minutie le détail des statistiques. Ceci explique que j'aie tenté de sérier le problème en présentant trois cas d'étude très différents, tant par les fluctuations spatiales et temporelles du magnétisme, que par leur échantillonnage par le bateau. Mais il n'est pas impossible qu'un jeu de données puisse mettre en péril le "moindre mal" que représente la géométrie $\mathcal{V}_3^x$.
La pratique sur sept campagnes océanographiques m'a montré qu'en procédant avec minutie, sept jours de travail sont nécessaires pour, partant de la bande de données issue du bateau, aboutir à une estimation acceptable des perturbations diurnes, lorsqu'elles sont oscillantes ... Certains peuvent juger que c'est long, mais quand on sait qu'une campagne océanographique peut durer plusieurs semaines, et que celle-ci peut se faire dans des conditions climatiques particulièrement difficiles (et là j'ai une pensée émue pour cette campagne Mohns qui m'a beaucoup marqué); quand de plus on prend conscience de la richesse de l'information qui est contenue dans ce magnétisme spatial; alors, fort de ces arguments, peut être conviendra-t-on que les données méritent une telle minutie. En tous cas, telle est mon opinion.

Le deuxième fait marquant dans cette approche du magnétisme spatio-temporel, c'est que finalement, si l'analyse est désastreuse (détection des fréquences difficiles) l'estimation spatiale est tout à fait acceptable et les cas d'étude Krakatao et Seaperc sont il me semble très démonstratifs à ce propos.
Ce contraste entre analyse et estimation signifie en premier lieu que nos techniques d'analyse sont inadaptées à ce phénomène et ceci, comme il vient d'être dit, en raison du voisinage de krigeage. Mais je souhaite ici attirer l'attention sur le fait suivant qui contribue à ce contraste entre détection et estimation : l'estimateur D*(t) des perturbations temporelles est fortement bruité, alors que l'estimateur spatial résultant Y*(x) est beaucoup moins erratique.
Qu'est-ce qui diffère entre les deux? Outre l'opération de soustraction $Y^*(x) = Z(x,t) - D^*(t)$ qui finalement décompose le bruit mono-dimensionnel de D*(t) dans le plan (ce qui est déjà une circonstance qui atténue le bruit sur la carte) il y a une opération de maillage et de cartographie de Y*(x). Or, ces opérations (et surtout le maillage) reviennent (une fois de plus ...) à effectuer une coupure sur le spectre de Y*(x), ce qui précisément permet de supprimer une grande partie du bruit que D*(t) a apporté à Y*(x).
Cette circonstance m'a permis d'obtenir des cartes acceptables, mais suscite quelques réflexions :

- Tout d'abord, une inquiétude : n'est-il pas discutable de confier à une opération de maillage, le soin de purifier un estimateur par trop bruité ? Ne faudrait-il pas mieux effectuer un pré-filtrage dans le temps de D*(t) Ceci paraît possible car l'estimateur à de bonnes chances d'être stationnaire (puisque périodique) de sorte que le concept de spectre est possible.
  De plus, il est en général connu à pas régulier (une minute) de sorte que l'opérateur TFD est utilisable pour effectuer cette coupure des hautes fréquences. Mais où effectuer cette coupure? Et ne risque-t-on pas de supprimer du détail essentiel, celui que la pratique glissante nous a permis d'estimer et qui fait que notre estimateur est (à juste titre) loin d'être sinusoïdal ?

C'est pourquoi je préconiserais une fois de plus et avant tout de supprimer l'effet sautant du voisinage. Après seulement, il sera possible de juger de l'opportunité d'une telle coupure spectrale.

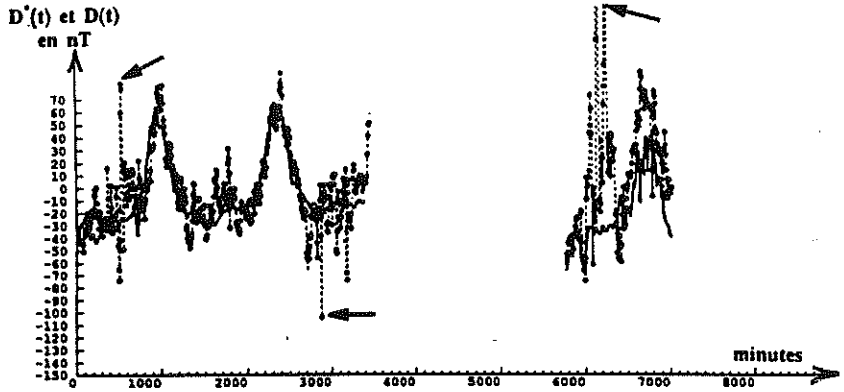
- La deuxième réflexion que me suscite l'effet de ce maillage (spatial) sur l'estimation, sort du cadre de cette thèse et concerne la cartographie d'une manière générale. L'impact que possède ce maillage n'est il pas inquiétant? Il se trouve que dans le cas des perturbations, ce maillage est bénéfique car, il semble qu'il filtre plus de bruit qu'il n'enlève de détail utile. Mais, d'une manière générale, qu'advient-il de la cartographie d'une quelconque variable ? Que perd-on exactement par ce maillage?
  On trouvera quelques éléments de réponse dans les lignes qui précèdent (recouvrement spectral, etc ...). Je n'ai pas d'autre réponse à apporter et je souhaitais simplement susciter une interrogation ... et une inquiétude...

Le troisième fait marquant dans ce travail me semble être le résultat des expériences de filtrage des sinusoïdes perturbatrices de l'image Gandhi (chapitre 5), non pas tant par les applications nouvelles qu'elles recèlent, mais surtout par le fait qu'elles complètent celles effectuées sur le magnétisme spatio-temporel.
Dans ces expériences, nous avons simplifié bon nombre de problèmes :

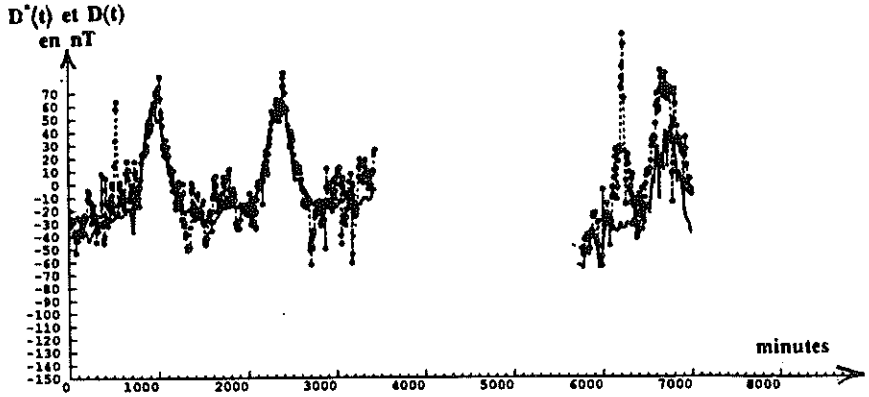
- Tout d'abord, nous n'avons effectué aucune analyse de l'image bruitée puisque lors de l'estimation, nous avons utilisé directement le variogramme de l'image non bruitée qui est inaccessible en pratique.
- La seconde simplification tient à la nature du bruit perturbateur qui n'est pas erratique mais strictement périodique puisque c'est une sinusoïde qui a été ajoutée à l'image.
- Une autre simplification par rapport au magnétisme, est celle que le modèle associé à l'image est celui d'une FAI-0, ce qui signifie que seule la fonction constante est utilisée en dérive (spatiale).
- Enfin, dernière simplification : comme l'image est à maille strictement régulière, et comme une translation dans le plan en est une également dans le temps (de balayage), une seule inversion du système linéaire est conduite. Les pondérateurs sont calculés une fois pour toute et l'estimateur temporel ne peut donc pas embarquer des fluctuations liées à de sporadiques problèmes d'inversion comme ce fut le cas pour le magnétisme spatio-temporel.

Le résultat de ces simplifications est que le filtrage a remarquablement agi, et il convient de savoir laquelle de ces simplifications s'est avérée la plus bénéfique (il ne fait aucun doute que toutes ont été bénéfiques ...).
Si, pour pouvoir répondre à cette question d'une manière définitive, il eût fallu reconduire ces estimations en modifiant chacun des paramètres ...(on peut penser notamment à bruiter l'image avec un signal très fluctuant comme par exemple un enregistrement de perturbations diurnes, disons Kakioka ou Seaperc ...) ... mon opinion est que le caractère très régulier du bruit n'est pas prépondérant sur le résultat. Pourquoi? Essentiellement parce que la phase et l'amplitude de la sinusoïde perturbatrice sont inconnues (d'une part) et que ces inconnues sont réestimées en chaque pixel de l'image (d'autre part) puisque l'estimation a été conduite en voisinage glissant. Si nous n'avions effectué cette estimation de phase et d'amplitude qu'une seule fois (disons dans une zone "calme " de l'image) et si les coefficients obtenus en un point avaient servi au filtrage partout, alors l'expérience eût été bien moins démonstrative. Mais le fait de reconduire l'estimation en tout point nous exposait à des fluctuations (celles des valeurs échantillonnées localement) qui auraient pu très bien nuire au filtrage, voire le détruire.
Or, qu'observe-t-on ? Un estimateur D*(t) très régulier et proche de la réalité, et ceci partout. Ce fait expérimental signifie qu'en chacun des 256x256 pixel, ce sont sensiblement les même phase et amplitude qui ont été obtenues. Ce résultat n'est-il pas remarquable quand on sait que le rapport signal sur bruit est approximativement de 1 pour cette image ?
Toujours en ce qui concerne la nature du bruit, on peut alors rétorquer que c'est la stricte identité entre le modèle (les cosinus et sinus en dérive) et la réalité (c'est bien une sinusoïde qui perturbe...) qui a permis ce résultat. A ceci il me faut mentionner certaines expériences mono-dimensionnelles ne figurant pas dans cette thèse, où nous tentions d'estimer des valeurs de perturbations par un Krigeage (Trigonométrique). Moyennant un ajustement du diamètre du voisinage, ces expériences montrèrent alors que l'on pouvait être aussi proche que souhaité des perturbations, et ceci seulement à l'aide d'une poignée de sinusoïdes comme fonctions de dérive. De plus, les variances expérimentales obtenues alors étaient de même ordre de grandeur que celles obtenues pour le même objectif (estimer D(t)) mais avec pour signal d'entrée le magnétisme spatio-temporel. Ceci signifie deux choses : d'abord que l'identité n'a pas à être stricte entre les fonctions du modèle et la "fonction réalité" (ce qui clôt le débat sur le bruit de l'image Gandhi) et ensuite que séparer ce qui appartient à l'espace de ce qui appartient au temps, est loin d'être un problème insurmontable, et à mon avis, c'est même une chose des plus aisée tant ces régionalisations (espace et temps) diffèrent l'une de l'autre.
En conséquence, il me semble que la simplification qui a apporté le plus de stabilité à l'estimation, ne tient pas dans la nature du bruit perturbateur, mais bien dans la nature de l'image, ou plus exactement dans la nature du modèle associé à cette image, à savoir sa

stationnarité.
Pour étayer mon propos, je souhaite mentionner le fait suivant : dans le chapitre 4 qui concerne l'estimation du magnétisme spatio-temporel , nous avons été amené à comparer le résultat de deux modèles (étiquetés quatre et cinq) qui ne différaient que par le degré k de la FAI, tous les autres paramètres (géométrie, covariance, etc...) étant égaux par ailleurs.
Lorsque k était égal à 1 , nous obtenions l'estimation suivante :



Lorsque k était forcé de prendre la valeur nulle, nous obtenions alors l'estimation suivante :



Que donne la comparaison de ces deux estimations ? Les explosions du premier cas disparaissent ou s'atténuent dans le second cas. Pourquoi ? Parce que les incontournables alignements de points ne rendent plus, pour k = 0, l'inversion impossible (ou instable), alors que lorsque k = 1, certaines estimations sont difficiles.
Dans l'image Gandhi, ces alignements de points dans le voisinage n'existent pas, et de toute façon le problème ne se pose pas puisque k = 0 ...
Rétrospectivement donc, il semblerait bien que la majorité des problèmes rencontrés pour mettre en oeuvre ce Krigeage spatio-temporel ne soient pas liés à son aspect trigonométrique

et temporel, mais surtout à l'aspect spatial du problème.
Ceci nous conduit alors à poser le problème très général de l'estimation et la cartographie (spatiale) des données sur des profils du type de ceux rencontrés en Géophysique marine : comment les traiter et les analyser ? Bien que débordant du cadre de cette thèse, ce problème mérite quelques remarques.
A mon avis, la solution passe par une géométrie de type $\mathcal{V}_3^x$: pour supprimer les alignements, il paraît évident d'étendre le voisinage dans toutes les directions du plan. S'il est vrai que cette extension tout azimut peut se faire sans aboutir nécessairement à une pseudo maille régulière du type $\mathcal{V}_3^x$, il semblerait que cette maille représente cependant une alternative acceptable à ce problème. Mais alors, n'est-il pas regrettable de ne pas utiliser l'information à des distances inférieures au pas de la maille ? Comment panacher la maille $\mathcal{V}_3^x$ et l'information en continu sur les profils ?

Je souhaiterais conclure cette discussion sur le filtrage des perturbations diurnes en magnétisme, par deux remarques portant sur la méthodologie.
Le premier point concerne le voisinage et sa géométrie que j'ai tenté de traiter comme un objet comparable à un degré de FAI ou à une covariance généralisée. Cet "objet voisinage", j'ai tenté de l'inférer comme les autres paramètres, ce qui m'a conduit à manipuler des ensembles de paramètres du type $\{ \mathcal{V}_3^x ; k = 1 ; T = 24^h ; K^1(h) \}$ pour bien montrer la manière dont j'ai appréhendé ce problème. Bien plus qu'une notation, il s'agit là d'une philosophie inhabituelle en Géostatistique, et ceci sans doute pour une raison très simple : outre le fait que peu de praticiens ont besoin de diverses géométries, la manipulation de tels objets est très lourde et notre manière de travailler (je n'ose dire nos programmes informatiques) ne sont pas adaptés à cette philosophie. Quand on sait l'incidence que possède la géométrie du voisinage sur le résultat, ne conviendrait-il pas de repenser notre manière de travailler en fonction de ce nouveau paramètre : le voisinage ? La tâche n'est pas aisée car si on a souvent parlé de diamètre et de pas d'échantillonnage, c'est bien souvent d'une manière abusive puisque même dans le cas (déjà marginal) des données sur profil, nous sommes quand même assez éloignés d'une maille régulière. Comment alors caractériser la géométrie de cet ensemble de points qui est si essentielle dans l'estimation ? Faut -il parler de densité d'échantillonnage ou de "maille en tendance" et établir des statistiques également sur les voisinages utilisés pour une succession d'estimations donnée ?
La deuxième remarque que je souhaitais faire concerne l'indice "prior" que j'ai parfois affecté à certains paramètres pour exprimer le fait que ce paramètre à une valeur arbitraire et non inférée. Il me semble que l'utilisation de cette indexation n'est pas anodine par les deux aspects nouveaux qu'elle révèle:

- D'une part, le fait même d'imposer à un modèle des paramètres que l'analyse n'a pas pu révéler (on pensera bien entendu aux périodes temporelles), ce qui est une chose assez inhabituelle en Géostatistique où la pratique consiste bien souvent à se contenter de ce que nous révèlent les données.
- D'autre part, le fait d'étiqueter cet a priori donne à nos modèles une nouvelle dimension d'analyse lorsqu'on les compare entre eux et qu'ils sont en concurrence pour un même jeu de données : leur degré "d'objectivité".

Dans l'exemple des perturbations diurnes, cet arbitraire a payé puisque même en allant à l'encontre de ce que nous révélait l'analyse ("pas de fréquences temporelles"), nous avons pu mener jusqu'à terme l'estimation. Or, qu'est ce qui nous a autorisé cet arbitraire sinon notre connaissance de la physique du phénomène ("on sait qu'il y a des fréquences temporelles sous cette latitude") de sorte que cet arbitraire n'en est pas un et que $T = T_{prior}$ seulement dans le système clos de nos tests. Ceci montre (mais était-ce vraiment à prouver ?) qu'en Géostatistique, être proche des données, c'est aussi être proche des lois physiques qui les régissent .

Sur la généralisation à d'autres PTP

La généralisation à des PTP autres que le magnétisme est loin d'être immédiate, sauf bien entendu si la configuration des données (et du problème) est comparable.
Ainsi, pour les fluctuations de la température à la surface de l'eau (cf chap. 1), lorsque l'échantillonnage est conduit par un bateau, on peut dire que le problème du filtrage des fluctuations diurnes dans le but de tracer des cartes spatiales de températures est homologue de celui du magnétisme.
De plus, si, toujours pour un phénomène spatio-temporel qui n'oscille que dans le temps, on dispose d'un véritable champ de séries temporelles pour lesquelles les ondulations sont en phase, on se trouve alors dans un cadre bien plus favorable que celui du magnétisme puisqu'on est à même de rompre cette fameuse et déjà citée "coplanarité" dans l'espace-temps de l'échantillonnage du bateau, en appliquant nos méthodes sur des voisinages cubiques (et non

plans).
Noter au passage que si dans cette thèse nous nous sommes focalisé sur des problèmes de filtrage, il est clair que les techniques élaborées sont utilisables pour une modélisation et une estimation du phénomène spatio-temporel lui-même Ainsi par exemple, au lieu d'extraire l'invariant Y(x) du magnétisme spatio-temporel Z(x,t), on peut très bien estimer Z(x,t) en des temps et des lieux inconnus. Pour ce faire, il n'est qu'à modifier le second membre d'une matrice. Mais une certaine prudence est nécessaire car (toujours pour le magnétisme), une estimation en un endroit différent des points informants ne peut être qu'une extrapolation puisqu'elle nous conduit à sortir du plan (dans l'espace-temps) des données !
En ce qui concerne l'exemple de la pollution (toujours dans le chapitre 1), il me semble que le problème de la modélisation de ce PTP est bien plus difficile puisqu'il paraît indispensable de régionaliser explicitement (et dans le temps) l'amplitude des sinusoïdes, si tant est qu'on ait choisi de modéliser les fluctuations temporelles de ce PTP par une série de Fourier. Cette nécessité est due au fait qu'on note une corrélation entre l'instant de la mesure et l'amplitude des ondulations (les fluctuations sont plus importantes l'hiver que l'été). Or, une régionalisation explicite de ces amplitudes nécessite d'utiliser des modèles de covariance extrêmement différents de ceux utilisés dans cette thèse, et très certainement d'une inférence et d'une mise en oeuvre des plus délicates.

De ce point de vue, les opérations faites sur le magnétisme peuvent être qualifiées de "frustes" puisque la manière dont nous avons régionalisé les amplitudes de nos sinusoïdes est des plus simplistes: nous avons travaillé localement et par voisinage glissant. C'était là un moyen bien commode d'éviter l'écueil d'une régionalisation explicite des amplitudes, tout en s'adaptant aux fluctuations temporelles locales de la réalité. Mais une telle simplicité a peu de chance de fonctionner pour des problèmes plus complexes comme ceux de la pollution.
Le dernier exemple de PTP sur lequel je souhaite m'étendre quelque peu est celui des polypiers, et ceci parce que cet exemple donne une autre dimension au PTP : celle que la période peut être un vecteur. Rappelons en effet que cette image représente la répétition dans le plan de sensiblement la même structure dite "élémentaire" (une cellule), de forme variable (mais pas trop ...) avec des implantation pas exactement prévisibles. Pour de tels phénomènes (notons les Z(x,y)) il peut s'avérer opportun de faire l'hypothèse explicitement que le champ est la répétition périodique d'une structure élémentaire (la cellule dans l'exemple) et de conduire l'analyse en conséquence, à savoir :

- D'analyser par TFD (ou variographie ou les deux) chaque structure élémentaire, et de calculer la moyenne des spectres (ou variogrammes) obtenus.
- D'analyser le champ complet globalement par TFD (ou variographie ou les deux).
- De comparer les caractéristiques globales et locales, dans le but d'établir un modèle du type : Z(x,y) = Dérive + répétition { structure élémentaire }.

En des termes plus imagés de Théorie du Signal, il s'agirait de dégager par analyse locale les trains d'ondes caractéristiques des structures répétées, puis de déposer ce train d'ondes sur des porteuses basse fréquence (le caractère répétitif) déterminées par l'analyse globale.

L'ensemble de ces remarques sur divers PTP nous conduit semble-t-il bien loin des expérimentations de cette thèse ? Il n'en est rien, car ces quatre exemples (magnétisme, température,pollution,polypiers) et la manière dont je propose de les traiter, reflètent ce qui est peut être le seul véritable message contenu dans cette thèse : celui d'accepter dans le modèle, ce caractère répétitif. J'ai toujours été frappé par l'attitude adoptée par bon nombre de praticiens lorsqu'ils sont confrontés à ce caractère répétitif : ils choisissent bien souvent une échelle de travail qui permet de négliger la tendance et ceci même si l'abondance des données permet de travailler à plus grande échelle. Or, un tel choix, s'il n'est pas justifié ("à l'échelle choisie, la tendance est elle vraiment négligeable ?") peut s'avérer désastreux sur le résultat.
Les opérations faites sur le magnétisme ne sont-elles pas la preuve qu'au lieu de refuser ce caractère répétitif il peut s'avérer utile de l'accepter en le modélisant (par exemple) par une série de Fourier ?
Ceci m'amène à reprendre l'image de la modulation dont il a déjà été fait mention dans le chapitre 1 : puisque Dame nature nous offre un signal modulé par son caractère répétitif, pourquoi ne pas procéder comme les milliards de récepteurs radiophoniques qui fonctionnent dans le monde : nous caler sur la porteuse, puis écouter ...

Sur la complémentarité spatial-spectral

Par ce titre je souhaite appuyer le fait que la vision d'un spectre et d'un variogramme sont des images complémentaires de la réalité. Si on reprend le fil de mon travail cette complémentarité n'est finalement manifeste que dans le chapitre 2 où ce sont des séries temporelles qui sont analysées et où l'on s'est évertué explicitement à séparer dans le spectre du PTP, ce qui peut être qualifié de "purement spectral" à savoir les fréquences qui émergent, puis de laisser l'autre partie du phénomène, sa composante non périodique (ou encore "purement spatiale") être modélisée par des techniques classiques en Géostatistique (utilisation de

variogrammes, etc...).
C'est lors de ces opérations qu'il a été fort utile d'utiliser la TFD pour calculer (pour la série Kakioka) les caractéristiques énergétiques exactes de ces fréquences émergentes. Mais ce qui paraît remarquable, c'est que dans le cas où (comme pour la série Huancayo) cette TFD n'est pas utile, l'ajustement manuel de ces caractéristiques spectrales s'est avéré tout aussi fructueux.
Tout ceci pour dire que ce caractère répétitif du PTP donne à ces fonctions cosinus qui servent pour l'ajustement de la partie périodique du variogramme, un statut à part par rapport aux autres modèles de variogramme, et ceci par leur interprétation énergétique.
Dans les chapitres 3 et 4 qui concernent le magnétisme spatio-temporel cette complémentarité est moins immédiate mais existe pourtant puisque l'extraction des oscillations temporelles du signal spatio-temporel s'est faite (comme pour une TFD) par l'intermédiaire de fonction trigonométrique. Mais cet aspect spectral de notre démarche diffère de celui du chapitre 2 puisque d'une part cette extraction d'énergie temporelle se fait sur des voisinages de diamètre bien souvent non résonnant avec les fréquences de l'analyse, et d'autre part que notre but est de capter bien plus de fréquences que celles spécifiées explicitement dans la série de Fourier associée (dans le modèle) à la dérive. On a en quelque sorte détourné la vocation initiale de l'analyse spectrale.
Mais par delà le cadre de cette thèse et du PTP, l'analyse de spectres, (simultanément de variogrammes) ne devrait-elle pas être plus fréquente dans notre pratique ?
Analyser aussi cette autre image du phénomène, complémentaire de celle des variogrammes, ne donnerait-elle pas une autre dimension à notre travail d'analyste?
Il me parait difficile d'aller plus loin que cette simple interrogation car c'est une pratique que je propose et les conséquences et résultats d'une telle pratique sont finalement imprévisibles. Peut-être que le fait de manipuler spectres et covariances expérimentales en regard de leurs modèles nous amènera à considérer d'autres modèles ou suscitera certaines interrogations sur nos pratiques. Qui sait ce que peut engendrer une pratique nouvelle ? Ne convient-il pas d'essayer tout simplement ?

Sur le krigeage de coefficients de Fourier

Si l'expérience de filtrage sur l'image à balayage peut être présentée comme complémentaire de celle du magnétisme, on pourrait juger que l'estimation des contours du visage (sur cette même image du chapitre 5) est quelque peu hors sujet, puisque Gandhi n'est pas un PTP. Et pourtant, les expériences faites sur le magnétisme, celles faites sur l'image à balayage, et l'estimation des contours du visage; ces trois ensembles d'expériences donc, possèdent en commun un objectif d'une extrême importance, et qui représente peut être la seule véritable originalité de mon travail : nous avons, tout au fil de ces expériences, effectué des Krigeages de coefficients de Fourier (voire de séries de Fourier).
Dans le cas du magnétisme (et de l'image à balayage), il s'agit de séries temporelles alors que dans le cas des contours, il s'agit d'une série de Fourier spatiale.
Le Krigeage de tels coefficients est une conséquence de la nature du PTP (certaines fréquences demandent à être estimées car elles émergent) et cette nécessité nous a permis d'envisager une extension méthodologique : celle du Krigeage de coefficients non émergents. C'est là me semble-t-il tout l'intérêt de cette expérience d'Analyse Krigeante (Trigonométrique) que de montrer qu'une telle extrapolation peut être possible et utile.
Mais en matière d'estimation de série de Fourier, il est une pratique très répandue : celle d'utiliser les TFD pour cela. Quel est le lien entre un coefficient estimé par Krigeage (Trigonométrique) et celui obtenu par TFD ? De même, nos tentatives de détection de fréquences temporelles dans le chapitre 3 peuvent-elles porter l'intitulé d'Analyse Harmonique? Nous allons tenter de répondre à ces deux questions.

Le lien avec l'Analyse Harmonique :

Par Analyse Harmonique d'un signal, nous entendons :

- Une détection des fréquences contenues dans le signal. et ceci à l'aide d'un balayage spectral de toutes les fréquences possibles, indépendamment des caractéristiques de la fenêtre d'analyse.
- Une évaluation des amplitudes (ou énergies) associées à ces fréquences.

Le KT tel que nous l'avons pratiqué consiste en une estimation d'amplitudes, et ceci pour des fréquences connues au préalable.
Nous ne réalisons donc que le deuxième point de l'Analyse Harmonique, puisque notre "balayage Spectral" s'est limité dans le meilleur des cas à deux fréquences.
En fait, nos outils se sont avérés encore plus pauvres que cela puisque nous avons eu toutes les peines du monde ne serait-ce qu'à confirmer l'existence dans le signal de fréquences connues par ailleurs. De plus, lorsque, comme pour Mohns, la fréquence ciblée par l'analyse n'est pas présente dans le signal, on n'assiste pas comme pour une Analyse Harmonique classique, à une simple absence d'énergie pour cette fréquence, mais dans notre cas à des instabilités numériques

(lors des estimations) et des incohérences de tests (lors des détections).
Si donc on compare nos techniques de détection à une Analyse Harmonique, force nous est de constater que nous n'avons fait que pratiquer une "Analyse Harmonique de pauvre".
Une circonstance atténuante tout de même : celle-ci fut pratiquée dans l'espace-temps et à maille non régulière ...

Le lien avec la Transformée de Fourier Discrète :

Par rapport à ce que nous appelons une Analyse Harmonique, la TFD présente un certain nombre de limitations :

- Les fréquences du balayage spectral ne sont pas quelconques mais celles associées à une série de Fourier.
- La période principale du balayage spectral est le diamètre de la fenêtre d'analyse.
- Les données doivent être à pas régulier.

Ces limitations nous rapprochent d'autant de nos techniques, et il paraît intéressant de se poser la question suivante: que représente un coefficient de Fourier obtenu par TFD par rapport au même coefficient (ie associé à la même fréquence) estimé par KT ?

Quelle que soit la dimension de la régionalisation, ainsi que celle des fréquences d'intérêt, il est immédiat de montrer que :

- si le KT est effectué en FAI-0 (ie pas de polynôme en dérive),
- si la covariance utilisée est un effet de pépite,
- si le diamètre de Krigeage est un multiple de la plus grande période spécifiée dans le système,

alors les coefficients de Fourier estimés par KT sont formellement identiques à ceux obtenus par TFD pour les mêmes diamètres et pas de la fenêtre d'analyse.

Comme la TFD n'est à ce jour vraiment opérationnelle qu'à deux dimensions, le KT peut donc être présenté comme une généralisation de celle-ci.

Cette généralisation est de 3 ordres :

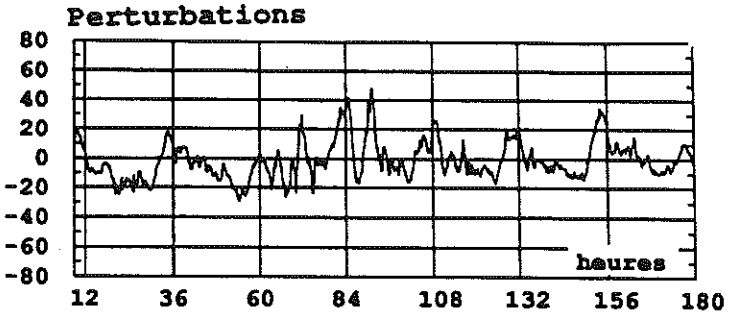
- le contexte de la régionalisation (espace-temps),
- le contexte de l'échantillonnage (non nécessairement périodique),
- le contexte du signal d'entrée (le bruit additif aux sinusoïdes n'est pas nécessairement blanc)

Comme le KT était utilisé pour estimer des perturbations qui ne se réduisaient pas à une sinusoïde mais à bien d'autres, nos estimations ont été conduites bien souvent sur des segments seulement de sinusoïdes, ce qui nous éloigne alors complètement des techniques spectrales puisque les fonctions du système cessent d'être des fonctions "d'analyse" pour devenir des fonctions de dérive. Mais pour un diamètre de voisinage multiple de la période ciblée, notre démarche s'apparente tout à fait à une TFD.

Si nous avons été amené à nous positionner par rapport à la TFD et à l'Analyse Harmonique, c'est parce que nous avons utilisé des fonctions de dérive trigonométrique. Or ce choix a été dicté par la nature même du PTP : son caractère répétitif.
Ce qui a rendu opérationelle l'utilisation de fonctions trigonométriques, c'est le fait que ces fonctions constituent une base, et qu'elles sont invariantes par translation. Il se pose alors une question bien simple : l'avenir ne va-t-il pas nous confronter à des phénomènes nécessitant d'autres bases fonctionnelles ? Je pense notamment aux fonctions de Walsh, aux sinus-cardinaux, ou tout simplement aux "Ondelettes" qui peuvent être présentées comme une généralisation des séries de Fourier.

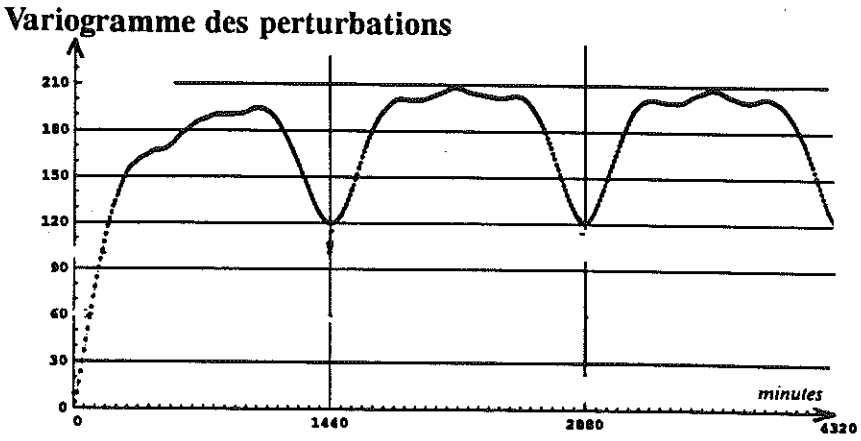
Sur le caractère esthétique et fascinant du PTP

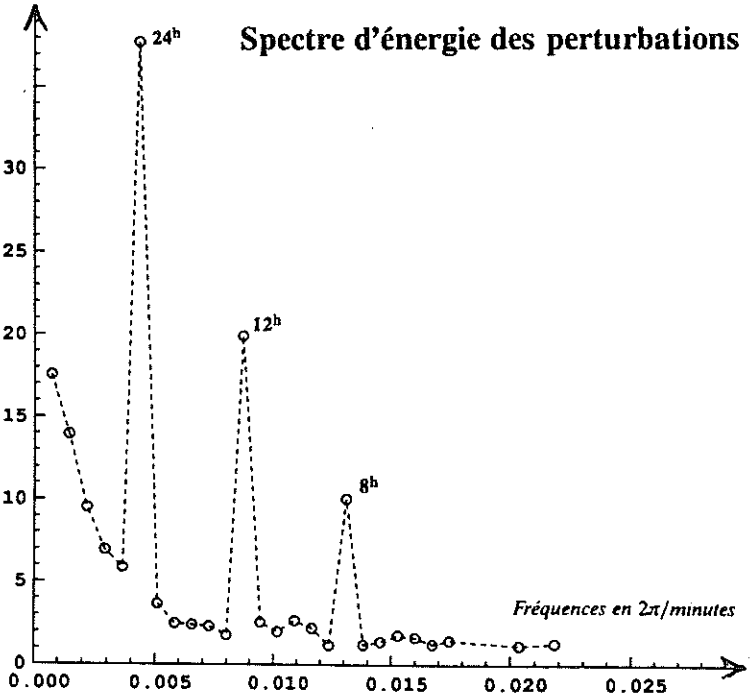
Je souhaite conclure ce travail par une tentative : celle de communiquer au lecteur la fascination qu'exerce sur moi le PTP, et plus particulièrement cette propriété qui le caractérise d'une "émergence" de certaines fréquences dans le spectre .
Dans ce but, je ne puis résister à l'envie de montrer à nouveau un extrait de 7 jours de la série temporelle Kakioka de perturbations diurnes qui a fait l'objet d'une analyse dans le chapitre 2.



Sur cette courbe, le caractère périodique du PTP est loin d'être manifeste ...

Puis de montrer le variogramme résultant de l'analyse sur 60 jours de mesure, ainsi que le spectre associé.





Et c'est là que la magie s'opère. En effet, comment ne pas être frappé par la régularité des ondulations dans ce variogramme; ou dans le spectre, par une émergence si brutale et manifeste des fréquences associées à 24^{h}, 12^{h} et 8^{h} ?
Tout ceci alors que le signal initial ne suggère que chaos et erraticité ...
On peut certes me rétorquer que la source de mon intérêt, c'est peut être tout simplement que certains phénomènes se soumettent parfois à une propriété très simple : la stationnarité, et que cette émergence n'est pas plus remarquable que celle de l'existence d'une espérance mathématique accessible par des moyennes spatiales sur des domaines suffisamment étendus...
Mais comment ne pas s'émerveiller devant la beauté, la régularité de ces ondulations ...
la manière frivole dont ce PTP accepte la stationnarité tout en refusant la permanence d'une variance ...vibrant hommage rendu à cette course éreintante de la terre autour d'elle même.

ANNEXES

A Outil d'analyse spectrale : la Transformée de Fourier Discrète page 179

B Outils Géostatistiques d'Analyse structurale page 183

C Krigeage(s) page 195

D Détail de la modélisation de la campagne Krakatao page 203

E Détail de la modélisation de la campagne Seaperc page 214

F Détail de la modélisation de la campagne Mohns page 226

A Outil d'analyse spectrale: la Transformée de Fourier Discrète

La Transformée de Fourier Discrète (TFD) est un opérateur qui existe pour les 2 raisons suivantes:

- La fonction z(t) (dont on souhaite réaliser une analyse spectrale) n'est connue qu'à travers son échantillonnage $z_e(t)$, qui n'est lui même connu qu'à travers une coupure $z_e(t,T)$ (ie $z_e(t)$ multipliée par une fonction rectangle de longueur T)
- On souhaite avoir un opérateur équivalent (en un sens à préciser) à la transformée de Fourier ($T\mathcal{F}$) d'une fonction continue, mais adapté à $z_e(t,T)$.

Nous allons construire l'opérateur TFD ainsi que son inverse (noté TFD^{-1}) et mentionner les contraintes liées à son utilisation.

La TFD (et son inverse) sont bien souvent mis en oeuvre à l'aide d'algorithmes rapides qu'on appelle " Transformée de Fourier Rapide " (TFR) encore appelés "Fast Fourier Transformer " (FFT) en Anglais. Ces algorithmes ne remettent pas en cause le principe de la TFD, et c'est pourquoi il n'en sera pas fait mention dans la suite. On pourra consulter **[TdS.2]** pour plus de détails sur la TFR.

A.1 Construction de la TFD

Position du problème :

Supposons que nous soyons en présence d'un ensemble infini $\{z(kT_e)\ , k \in \mathbb{Z}\}$ de valeurs qu'on supposera informées à pas régulier T_e, et appelons $z_e(kT_e)$ cette fonction. Nous allons pour l'instant oublier que $z_e(kT_e)$ peut être la version échantillonnée d'une fonction continue z(t).

L'intégration qui exprime la $T\mathcal{F}$ d'une fonction s'écrit dans ce cas discret:

$$X_e(f) = \sum_{k=-\infty}^{+\infty} z_e(kT_e)\ e^{-j2\pi\frac{f}{f_e}k} \qquad \text{avec } f_e = \frac{1}{T_e} \tag{a.1}$$

Comme $X_e(f)$ est périodique, ie:

$$X_e(f + f_e) = X_e(f) \tag{a.2}$$

la formule a.1 s'identifie au développement en série de Fourier ($S\mathcal{F}$) de $X_e(f)$, dont les coefficients $z_e(kT_e)$ peuvent être calculés sur une seule période de longueur f_e :

$$z_e(kT_e) = \int_{-\frac{f_e}{2}}^{+\frac{f_e}{2}} X_e(f)\ e^{j2\pi\frac{f}{f_e}k} df \tag{a.3}$$

Discrétisation des fréquences :

Comme dans a.3, f est une variable continue, la mise en oeuvre pratique de cette formule est impossible. On adopte donc la discrétisation suivante:

$$\Delta f = \frac{f_e}{N} \tag{a.4}$$

"N" étant pour l'instant un entier arbitrairement choisi.

Cette discrétisation permet d'approximer l'intégration a.3 par:

$$z_e(kT_e) \simeq \overline{z}_e(kT_e) \tag{a.5}$$

avec :

$$\overline{z}_e(kT_e) = \Delta f \sum_{n=-\frac{N}{2}}^{-1+\frac{N}{2}} X_e(n\Delta f) e^{j2\pi\frac{n}{N}k} \tag{a.6}$$

Conséquences de la discrétisation :

La fonction $\overline{z}_e(kT_e)$ donnée par a.6 est périodique de période N. On montre de plus que:

$$\overline{z}_e(kT_e) = f_e \; rep_N[z_e(kT_e)] \qquad (a.7)$$

ie que $\overline{z}_e(kT_e)$ est la répétition périodique de $z_e(kT_e)$.

La formule a.7 signifie que la discrétisation des fréquences entraîne une répétition dans l'espace de la fonction, avec tous les problèmes de recouvrement (dans l'espace physique et non comme usuellement dans l'espace des fréquences) que cela implique.

Si la fonction $z_e(t)$ est nulle hors de $[-N\frac{T_e}{2}, +N\frac{T_e}{2}]$, alors l'approximation a.5 est exacte pour k prenant les valeurs $\frac{-N}{2}, -\frac{N}{2}+1, \ldots\ldots, -1+\frac{N}{2}$

Si tel n'est pas le cas, cette discrétisation entraîne un phénomène de recouvrement dans le temps.

Coupure de champ :

Une deuxième hypothèse est maintenant nécessaire: celle que $z_e(t)$ est nulle hors d'un champ T, et on pose alors:

$$N = \frac{T}{T_e} \qquad (a.8)$$

Si tel est le cas, il y a identification entre $z_e(kT_e)$ et sa restriction à T : $z_e(kT_e,T)$.
La formule inverse a.1 se réduit alors à une sommation finie:

$$X_e(n\Delta f) = \sum_{k=-N}^{-1+N} z_e(kT_e)\; e^{-j2\pi\frac{n}{N}k} \qquad (a.9)$$

Définition :

Les formules a.6 et a.9 définissent respectivement les TFD directes et inverses sur un champ T (centré en $\emptyset$) d'une fonction $z_e(kT_e)$.

A.2 <u>En résumé</u>

Les TFD directes et inverses ne s'identifient donc qu'aux $T\mathcal{F}$ directes et inverses de $z_e(kT_e,T)$. Si on souhaite évaluer la $T\mathcal{F}$ de $z_e(kT_e)$ à partir de la TFD de $z_e(kT_e,T)$, la qualité de l'approximation sera directement liée aux dimensions de T. Si de plus on souhaite évaluer la $T\mathcal{F}$ de z (la fonction continue dont $z_e(kT_e)$ est l'échantillonnage) à partir de cette TFD, aux approximations précédentes s'ajoutent celles des distorsions liées à l'échantillonnage.
Comme bien souvent on désigne par " TFD[z]" ce qui n'est en fait qu'une "$T\mathcal{F}$ [$z_e(kT_e,T)$] ", nous dirons pour résumer que ce qui sépare TFD[z] de $T\mathcal{F}$ [z] , c'est:

- le recouvrement spectral liée au passage: $z(t) \rightarrow z_e(kT_e)$,
- les erreurs d'approximations et de recouvrement spatial liées au passage: $z_e(kT_e) \rightarrow \overline{z}_e(kT_e)$,
- les erreurs de coupure de champs liées au passage: $\overline{z}_e(kT_e) \rightarrow \overline{z}_e(kT_e, T)$.

Cette première analyse supposait que l'on souhaitait atteindre la $T\mathcal{F}$ [z] à l'aide de sa TFD. Maintenant, on peut très bien choisir d'analyser $z_e(kT_e,T)$, et dans ce cas, on a :

$$TFD[z_e(kT_e, T)] = T\mathcal{F}[z_e(kT_e, T)] \text{ pour } k \in N \text{ et tel que } kT_e \in [-\frac{T}{2}, +\frac{T}{2}] , \qquad (a.10)$$

$$= rep_T[T\mathcal{F}[z_e(kT_e, T)]] , \qquad \text{hors du champ.}$$

$$= \text{rep}_T[T\mathcal{F}[z_e(kT_e, T)]], \qquad \text{hors du champ.}$$

Donc, lorsqu'on analyse une suite finie d'échantillons par TFD, on obtient une périodisation de la suite finie de la $T\mathcal{F}$ de ces échantillons, et la TFD^{-1} reconstitue ceux-ci sur le domaine de TFD.

A.3 Application aux fonctions périodiques

Si $z_e(kT_e,)$ est périodique de période T, alors l'égalité a.10 est vraie $\forall$ k, et sa TFD donne (à un coefficient multiplicatif Δf près) les coefficients de Fourier associés. C'est pourquoi dans ce cas, il convient d'analyser la fonction sur un domaine de taille multiple de T, pour que les fréquences harmoniques de la TFD soient cohérentes avec celles contenues dans $z_e(kT_e,)$. Si tel n'est pas le cas, l'énergie contenue dans $z_e(kT_e)$ se répartira sur des fréquences discrètes différentes de la réalité : on dit alors qu'il y a diffusion de l'énergie. Ce phénomène entraîne une interprétation erronée du spectre.

A.4 Application à l'interpolation

Une autre manière de présenter la TFD, consiste à la définir comme un opérateur qui à N valeurs discrètes permet d'associer N fréquences et N couples (A_i, B_i) de coefficients de Fourier (ceux associés à a.9), de sorte que la reconstitution:

$$z_{TFD}(t) = a_\beta + \sum_{n=1}^{N} [\, A_n \cos(2\pi \frac{n}{N} t) + B_n \sin(2\pi \frac{n}{N} t) \,] \qquad (a.11)$$

soit exacte aux points $t = kT_e$, ie :

$$z_{TFD}(kT_e) = z(kT_e) \qquad \text{pour k} : \frac{-N}{2}, -\frac{N}{2}+1, \ldots\ldots, -1+\frac{N}{2} \qquad (a.12)$$

A.5 Généralisation à 2 dimensions

Cet opérateur TFD se généralise à 2 dimensions.
La TFD directe d'une fonction z(x,y) est donnée par:

$$X_e(n\Delta f_x, m\Delta f_y) = \sum_{k=\beta}^{-1+N_X} \sum_{l=\beta}^{-1+N_y} z_e(kT_x, lT_y)\ e^{-j2\pi(\frac{n}{N_x}k + \frac{m}{N_y}l} \qquad (a.13)$$

et la TFD^{-1} par

$$z_e(kT_x, , lT_y) = f_x f_y \sum_{n=-\frac{N_x}{2}}^{-1+\frac{N_x}{2}} \sum_{m=-\frac{N_y}{2}}^{-1+\frac{N_y}{2}} X_e(n\Delta f_x, m\Delta f_y)\ e^{j2\pi\,(\frac{n}{N_x}k + \frac{m}{N_y}l)} \qquad (a.14)$$

avec :

$N = (N_x, N_y)$: vecteur nombre d'échantillons en x et y

$T = (T_x, T_y)$: vecteur pas d'échantillonnage en x et y

$f = (f_x, f_y) = (\frac{1}{T_x}, \frac{1}{T_y})$: vecteur des fréquences d'échantillonnage

Les mêmes précautions d'emploi que dans le cadre mono-dimensionnel sont à respecter.

A.6 Notations

Dans cette thèse, lorsque nous analysons une fonction par TFD, nous désignons par :

D_{TFD} : le vecteur diamètre du champ d'analyse, (a.15)

p_{TFD} : le vecteur pas d'échantillonnage de la fonction. (a.16)

B Outils géostatistiques d'analyse structurale

Nous présentons très succinctement dans ce qui suit les fonctions variogramme et covariance généralisée.
Les problèmes d'ergodicité ne seront pas abordés et on consultera **[GST-5]** et **[GST-10]** à leur propos.

L'existence de chacune de ces fonctions dépend d'hypothèses particulières sur une VR que nous appellerons z(x), x étant le vecteur des coordonnées.

La première étape, commune à ces deux fonctions, consiste à interpréter z(x) comme la réalisation d'une Fonction Aléatoire (FA) Z(x). C'est sur cette dernière que porteront ces hypothèses particulières.

Une fois ces fonctions définies à partir de la FA Z(x), nous expliquerons comment le Géostatisticien effectue leur inférence à partir de la réalisation unique z(x) dont il dispose.

B.1 La fonction variogramme $\gamma(.)$ (réf [GST-4])

Définition :

On suppose que la FA Z(x) est intrinsèque d'ordre 2, c'est à dire:

$$E\ (\ Z(x+h) - Z(x)\ ) = \emptyset \qquad \forall\ x \text{ et } h \tag{b.1}$$

$$E\ (\ Z(x+h) - Z(x)\ )^2 = 2\ \gamma(h) \qquad \forall\ x \text{ et } h \tag{b.2}$$

La formule b.2 définit la fonction variogramme notée $\gamma(h)$.

Propriétés :

Ses propriétés sont d'être paire, nulle à l'origine, de croître moins vite que h^2 et d'être négative conditionnelle, c'est à dire que pour tout ensemble $\{\lambda^\alpha\}$ de coefficients tels que:

$$\sum_\alpha \lambda^\alpha = \emptyset \tag{b.3}$$

alors

$$\mathrm{Var}\Big(\ \sum_\alpha \lambda^\alpha Z(x_\alpha)\ \Big) = -\sum_\alpha \sum_\beta \lambda^\alpha \lambda^\beta \gamma(|x_\alpha - x_\beta|) \quad \geq \emptyset \tag{b.4}$$

Si $\gamma(h)$ est borné, on en déduit que Z(x) est une FA stationnaire d'ordre 2 (FAST) ayant une covariance C(h).

Si de plus $\gamma(\infty)$ existe, on peut exprimer cette covariance comme suit :

$$C(h) = \gamma(\infty) - \gamma(h) \qquad \forall\ h \tag{b.5}$$

avec :

$$\gamma(\infty) = C(\emptyset) \tag{b.6}$$

Le variogramme est donc un outil plus général que la covariance (centrée) puisqu'il permet l'inférence de celle-ci en évitant l'écueil de l'estimation de E(Z).
De plus, il permet de modéliser une classe de phénomènes plus grande que les FAST.

Exemples :

Comme exemples de variogrammes, on peut citer:

- Le modèle **PÉPITIQUE:**

$$\gamma_{pep}(h) = c\ (1 - pépite(h)) \qquad \text{où pépite(h) est une fonction nulle partout sauf en 0 où elle égale 1.} \qquad (b.7)$$

- Le modèle **SPHÉRIQUE:**

$$\gamma_{sphé}(h) = c\ \left(\frac{3}{2}\frac{|h|}{a} - \frac{1}{2}\frac{|h|^3}{a^3}\right) \qquad \forall\ |h| \leq a, \text{ égal à c ailleurs} \qquad (b.8)$$

Comme $\gamma(\infty)$ existe, il est possible d'associer une covariance $C_{sphé}(h)$, à $\gamma_{sphé}(h)$ par la formule b.5

Cette fonction d'auto–corrélation sphérique est bien connue des Théoriciens du Signal puisqu'elle représente la variabilité des fluctuations du courant dans une diode (à tube) à température limitée (plus exactement une "température–limited–parallel plane diode" en anglais).
Ce bruit de percussion ("Shot noise" en Anglais) est dû à l'émission imprévisible d'électrons, par la cathode de la diode, et aux chocs des électrons entre eux.

Pour plus de détail, on consultera RACK, 1938, (réf [TdS–4])

- Le modèle **PRISMATO–MAGNETIQUE:**

$$\gamma_{pmag}(h) = c\ \left(1 - \frac{1}{(1 + (\frac{h}{a})^2)^{\frac{3}{2}}}\right) \qquad (b.9)$$

- Le modèle à **EFFET DE TROU** (cas mono–dimensionnel) :

$$\gamma_{Trou}(h) = c\ (1 - e^{-a|h|}\ \cos(2\pi f_0 h)) \qquad (b.10)$$

- Le modèle **PÉRIODIQUE** (période T):

$$\gamma_{pério}(h) = \sum_{i=1}^{Nfreq} \xi_i\ (1 - \cos(\frac{2\pi}{T_i}h)) \qquad \xi_i \geq 0 \qquad (b.11)$$

- Le modèle **PUISSANCE:**

$$\gamma_{puis}(h) = |h|^{\lambda} \qquad \lambda < 2 \qquad (b.12)$$

Remarques :

Lorsque $\gamma(\infty)$ existe, on dit que $\gamma(h)$ est à "palier". C'est le cas des modèles $\gamma_{pep}(.)$, $\gamma_{sphé}(.)$, $\gamma_{pmag}(.)$, $\gamma_{trou}(.)$ (dans ces exemples, le palier a pour valeur "c"). Ce palier s'interprète comme la variance de la FA Z(x) associée à $\gamma(.)$.

Lorsqu'il existe une valeur (appelée "portée") telle que pour $|h|$ plus grand que cette valeur, $\gamma(h)$ atteint son palier, on dira que le modèle est à portée finie . C'est le cas des modèles $\gamma_{pep}(.)$ (portée nulle) et $\gamma_{sphé}(.)$ (portée de valeur "a" dans la formule b.8). L'existence d'une portée "a" signifie l'absence de corrélations spatiales pour des distances supérieures à "a".

Bien que le modèle $\gamma_{trou}(.)$ soit à portée infinie, on parle cependant dans ce cas (et dans les cas similaires) de "portée pratique": la distance au delà de laquelle $\gamma(.)$ atteint 95% de son palier.

Particularités du modèle périodique :

Parmi les modèles présentés, le modèle $\gamma_{\text{pério}}(.)$ possède les particularités suivantes:

- Il est borné par la quantité $\sum_{i=1}^{Nfreq} \xi_i$, mais $\gamma(\infty)$ n'est pas défini.
- On peut cependant lui associer une variance égale à $\sum_{i=1}^{Nfreq} \xi_i$, et donc une covariance.
- Il n'y a jamais extinction de la corrélation spatiale.
- A l'inverse des autres modèles (sauf $\gamma_{\text{puis}}(.)$ dont la négativité est conditionnelle), il n'est pas strictement défini négatif, ie $\text{Var}(\sum_{\alpha} \lambda^{\alpha} Z(x_{\alpha})) = \emptyset$ n'entraîne pas λ^{α} tous nuls.

B.2 Inférence du variogramme

En général, nous ne connaissons qu'une réalisation discrète $\{z(x_{\alpha})\}$ sur un champ limité de la FA Z(x) et c'est pourquoi on calcule préliminairement le variogramme expérimental $\gamma^*(h)$, donné par:

$$\gamma^*(h) = \frac{1}{N(h)} \sum_{N(h)} (z(x_{\alpha}) - z(x_{\beta}))^2 \qquad (b.13)$$

où N(h) représente le nombre de couples de points de mesure $(z(x_{\alpha}), z(x_{\beta}))$ distants de h.

Pour que, à h fixé, la formule b.13 soit statistiquement représentative, il est nécessaire que N(h) soit grand. Ceci implique que bien souvent, $\gamma^*(.)$ est calculé une fois pour toute sur le champ complet.

Suite à ces calculs et lorsque cela est possible, on essaye de trouver un modèle de variogramme qui soit cohérent avec $\gamma^*(.)$. C'est la phase d'ajustement de la modélisation. Elle s'effectue bien souvent visuellement à l'aide d'informatique graphique, et en utilisant des modèles dits "autorisés" : des fonctions variogrammes telles que données par les formules b.7 à b.12 et dont on cherche une combinaison linéaire (à coefficients positifs) qui passe "au mieux" par les points de $\gamma^*(.)$.

Si cet ajustement peut se faire uniquement à l'aide de modèles stationnaires (d'ordre 2), on en déduit la possibilité de considérer z(x) comme la réalisation d'une FAST, dont la covariance est donnée par la formule d'échange b.5.

B.3 La fonction covariance généralisée K(h)

Avertissement :

La définition qui suit est, par pur souci de simplicité, extrêmement informelle. On consultera donc Matheron (réf [GST-7]) pour la définition exacte.

Définition :

On dira qu'une FA Z(x) est une Fonction Aléatoire Intrinsèque d'ordre k (FAI-k), si elle appartient à une classe de FA équivalentes en ceci que:

pour tout ensemble $\{\lambda_{\alpha}\}$ de coefficients tels que:

$$\sum_{\alpha} \lambda^{\alpha} f_{\ell}(x_{\alpha}) = \emptyset \qquad \forall\ \ell \leq k \qquad (b.14)$$

alors:

$$E\left(\sum_{\alpha} \lambda^{\alpha} Z(x_{\alpha})\right) = \emptyset \qquad (b.15)$$

et:

$$Var\left(\sum_{\alpha} \lambda^{\alpha} Z(x_{\alpha})\right) = \sum_{\alpha} \sum_{\beta} \lambda^{\alpha} \lambda^{\beta} K_{\alpha\beta} \qquad (b.16)$$

La formule b.16 définit la fonction Covariance Généralisée (CG) K(h) (l'expression " $K_{\alpha\beta}$ " y représente $K(|x_{\alpha} - x_{\beta}|)$).

Dans b.14, les fonctions $f_{\ell}(x)$ sont des fonctions polynomiales en x de degré l. On désignera par Λ_k l'ensemble des ensembles $[\lambda_{\alpha}]$ de coefficients vérifiant b.14 et on appellera Combinaisons Linéaires Autorisées d'ordre k (CLA-k) toute combinaison linéaire $\sum_{\alpha} \lambda^{\alpha} Z(x_{\alpha})$ de tels $[\lambda_{\alpha}]$.

Dans la suite, on ne regardera plus Z(.) qu'à travers de telles CLA-k.

Propriétés et exemples :

Les propriétés de K(h) sont d'être paire et positive conditionnelle, c'est-à-dire que pour toute CLA-k $[\lambda_{\alpha}]$, l'expression b.16 est non négative.

Comme exemple de CG, on peut citer:

- le modèle **SPLINE**:

$$K(h) = b|h|^{2} \log|h| \qquad b \geq \emptyset,\ k \geq 1 \qquad (b.17)$$

- le modèle **POLYNOMIAL**:

$$K(h) = -b_{\emptyset}\text{pépite}(h) + \sum_{p=\emptyset}^{k} (-1)^{p+1}\, b_{2p+1}\, |h|^{2p+1} \qquad b_p \geq \emptyset \qquad (b.18)$$

Remarques :

Une Fonction Aléatoire d'ordre 2 est une FAI-$\emptyset$ pour laquelle Λ_k est l'ensemble des CLA-k qui ne filtrent que les constantes (k= $\emptyset$), et on a:

$$K(h) = -\gamma(h) \qquad (Z(x)\ \text{FAI-}\emptyset) \qquad (b.19)$$

Les CLA-$\emptyset$ associées sont donc les accroissements de la FA.

De même, une FAST peut être présentée comme un cas particulier de FAI-k pour laquelle Λ_k contient toute combinaison linéaire, sans contrainte aucune. On a alors:

$$K(h) = C(h) \qquad (Z(x)\ \text{FAST}) \qquad (b.20)$$

Lorsque les fonctions $f_{\ell}(x)$ de la formule b.14 sont des fonctions de la forme:

$$f(x) = P(x)\ e^{<\omega\ ,\ x>} \qquad (b.21)$$

avec :

$P(x)$: un polynôme en x,

$x = (x_1, \ldots, x_n)$: le vecteur des coordonnées

n : la dimension de l'espace,

$\omega = (\omega_1, \ldots, \omega_n)$: un vecteur de pulsations complexes,

$$< \omega\ , x > = \sum_{i=1}^{n} \omega_i\ x_i \quad ,$$

on parlera alors de Fonctions Aléatoires Intrinsèques Généralisées (FAIG ou FAI-$\mathcal{F}$) sur l'espace $\mathcal{F}$ engendré par les fonctions f(x). On montre que si $\{\lambda_\alpha\}$ vérifie b.14, avec des fonctions de la forme b.21, alors la formule b.15 est vraie et il existe une fonction K(h) telle b.16 est vraie.

Comme exemple de covariance généralisée associée à une FAI-$\mathcal{F}$,on peut citer (réf [GST-11] et [GST-2]) toute combinaison linéaire (à coefficients positifs) de fonctions de la forme :

$$K(h) = -\cos(2\pi < f, h >)\ \gamma(h) \qquad \text{(b.22)}$$

avec :

$f = (f_1, \ldots, f_n)$: un vecteur de fréquences réelles,

$\gamma(h)$: une fonction variogramme.

Dans ce cas, l'espace $\mathcal{F}$ est celui engendré par les 2 fonctions complémentaires $\cos(2\pi < \omega, x >)$ et $\sin(2\pi < \omega, x >)$.

On peut encore citer:

$$K(h) = -\cos(2\pi < f_1, h >)\cos(2\pi < f_2, h >)\ \gamma(h) \qquad \text{(b.23)}$$

fonction pour laquelle l'espace $\mathcal{F}$ est celui engendré par les cosinus et sinus des vecteurs fréquences ω_1 , ω_2 , $(\omega_1 + \omega_2)$ et $(\omega_1 - \omega_2)$.

B.4 Inférence des caractéristiques de la FAI-k

Telle que nous la pratiquons, l'inférence de ce modèle est indissociable de l'utilisation que nous en ferons. Dans notre cas, il s'agira de l'estimation de la VR z(x) (ou d'une partie de celle-ci) à l'aide d'un opérateur que nous nous contenterons pour l'instant d'appeler "estimateur" et qui dans les faits est le Krigeage (cf annexe C). Les procédures décrites sont celles du logiciel BLUEPACK (réf [GST-15]).

Le voisinage $\mathcal{V}$:

Cet "estimateur" s'appuyant sur tout ou partie des échantillons $\{z(x_\alpha)\}$ contenus dans le champ, il se pose le problème du choix de l'information à utiliser pour le mettre en oeuvre.

Nous appelons "voisinage de l'estimateur" (noté $\mathcal{V}$ ou encore $\mathcal{V}(x)$) l'ensemble des points $\{z(x_\alpha)\}$ permettant l'estimation en un point x.

Nous disons que $\mathcal{V}$ est "unique" lorsqu'il contient tous les points du champ, et "glissant" lorsqu'il ne contient qu'une infime partie de celui-ci et se déplace sur le champ.

La mise en oeuvre de l'estimateur nécessitant un système numérique de taille liée à celle de $\mathcal{V}$, nous n'utilisons dans cette thèse qu'un $\mathcal{V}$ glissant, afin de ne pas alourdir les calculs informatiques.

Une autre raison de ce choix est qu'un $\mathcal{V}$ glissant permet d'adapter le modèle aux fluctuations locales de la VR .

Comme le modèle et son utilisation sont liés, $\mathcal{V}$ représente également la fenêtre d'analyse de la VR z(x), ainsi que le domaine de validité des caractéristiques de la FA inférées sur ce domaine.

Lorsque cela s'avère utile, on désigne par " $\mathcal{D}$ " le vecteur diamètre de $\mathcal{V}$, et par " $\mathcal{P}$ " son vecteur pas d'échantillonnage. On désigne par "Npoint" le nombre de points informants contenus dans $\mathcal{V}$.

Principe de l'inférence :

Il consiste en une succession "d'essais-erreurs" sur les paramètres "k" et "K(h)", et c'est pourquoi nous parlons de "boucles-k" (inférence du degré k) et de "boucles-K" (inférence de la CG K(h)).

A ces boucles s'ajoutent celles d'estimations "in-situ", où le modèle est directement testé sur les données. On parle alors de "boucles- K^* " (validation du modèle).

Dans chacune de ces boucles, nous construisons des CLA-k sur lesquelles nous confrontons hypothèse et réalité (figure b.1)

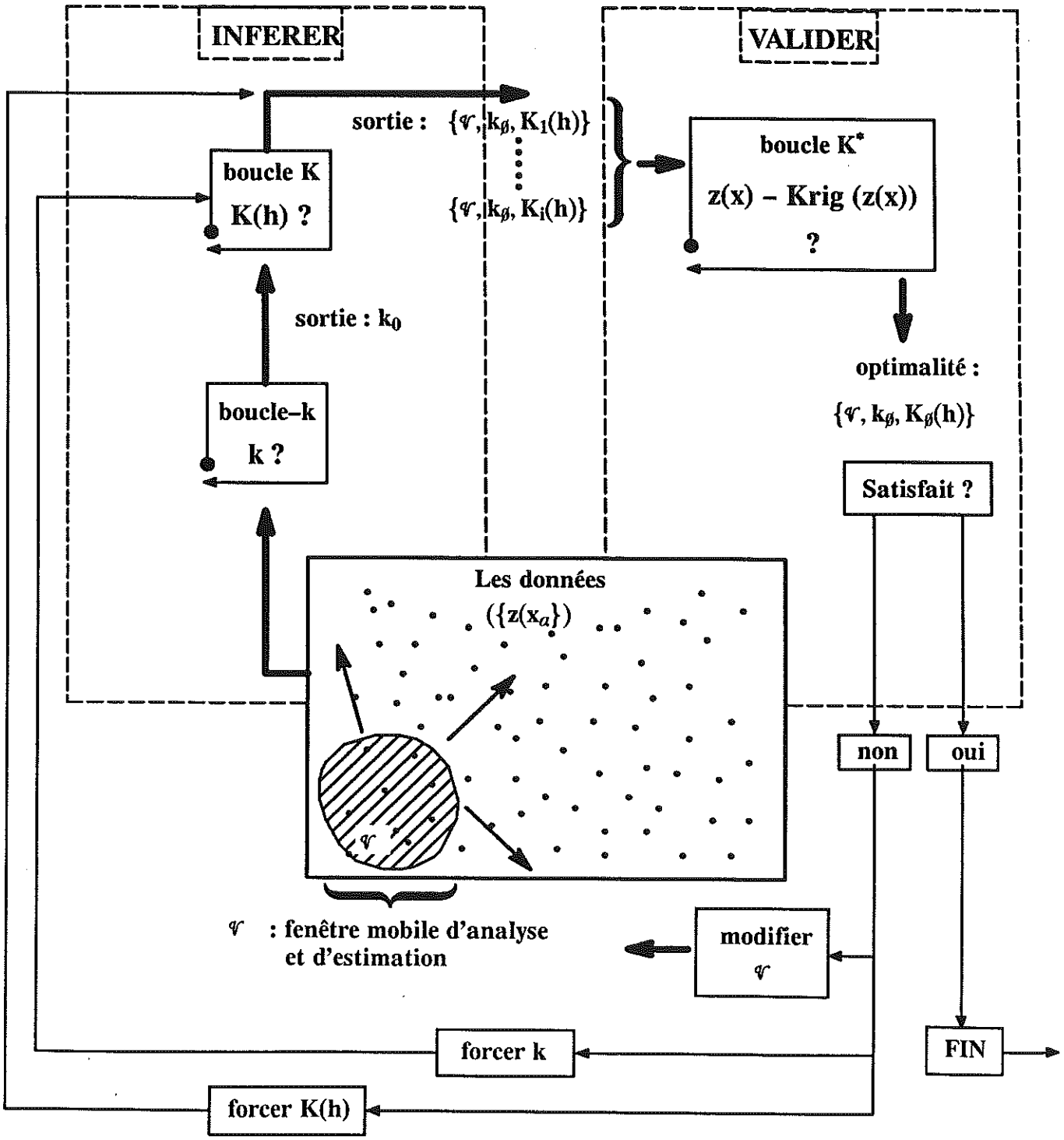


Figure b .1 : Inférence des caractéristiques de la FAI-k

Choix de zones tests :

Ayant adopté le principe d'un $\mathcal{V}$ glissant, et souhaitant obtenir un estimateur optimal (en un sens précisé ultérieurement) en moyenne sur le champ complet, la première étape consiste à choisir le nombre "nvois" de $\mathcal{V}$ sur lesquels s'effectuera l'inférence, et à définir l'ensemble $\{(x_1, \ldots, x_{nvois})\}$ des centres de ces $\mathcal{V}(x_i)$, de sorte qu'ils soient répartis d'une manière homogène sur le champ.

C'est sur chacune de ces "zones tests" $\mathcal{V}(x_i)$ que le paramètre k sera calculé; et celui retenu finalement sera celui qui fournira le meilleur estimateur sur l'ensemble de la zone.

Inférence de k (boucle-k) :

Fixons l'une de ces zones test (étiquetée "ivois") , faisons l'hypothèse que $k = k_{prior}$ (ie "k", valeur à priori) et procédons par étape (figure b.2) :

- **La première étape** consiste à étendre spatialement chaque $\mathcal{V}(x_i)$ de sorte que sa population passe de Npoint à 2Npoint. Nous appellerons $[\mathcal{V}(x_i)]_{étendu}$ ce nouveau voisinage. L'objectif de cette extension est de doubler la population.
- **La deuxième étape** consiste en le partitionnement de $[\mathcal{V}(x_i)]_{étendu}$ en deux ensembles $[\mathcal{V}(x_i)]_{R_1}$ et $[\mathcal{V}(x_i)]_{R_2}$. Cette décomposition s'effectue en étiquetant par un entier compris entre 1 et 2Npoint et par distance croissante par rapport au point ciblé par l'analyse, chacun des éléments de $[\mathcal{V}(x_i)]_{étendu}$ et en affectant à $[\mathcal{V}(x_i)]_{R_1}$ les éléments de rangs impairs et $[\mathcal{V}(x_i)]_{R_2}$ les éléments de rangs pairs. On espère, par cette opération, augmenter les effectifs statistiques, sans nuire à la géométrie de $\mathcal{V}$ fixée initialement, tout en détruisant la géométrie de l'échantillonnage (du bateau) qui peut altérer la perception de k.
- **La troisième étape** consiste, pour chaque point de $[\mathcal{V}(x_i)]_{R_1}$ (resp. $[\mathcal{V}(x_i)]_{R_2}$) à effectuer son estimation $z^*(x_\beta)$ en minimisant la forme quadratique suivante (ajustement par moindre carré) :

$$\sum_{\{x_\alpha\}} [\ P_{\alpha\,\ell}(k_{prior}) - z(x_\alpha)\ ]^2 \qquad (b.24)$$

avec :

$$P_{\alpha\,\ell}(k_{prior}) = \sum_{\ell=\beta}^{k_{prior}} a_\ell f_\ell(x_\alpha) \qquad (b.25)$$

où $\{x_\alpha\}$ est l'ensemble des points de $[\mathcal{V}(x_i)]_{R_1}$ (resp. $[\mathcal{V}(x_i)]_{R_2}$), $f_\ell(x_\alpha)$ un polynôme de degré l en x_α et $\{a_\ell\}$ l'ensemble des coefficients à déterminer.

Ces opérations permettent de dégager autant d'erreurs $(z(x_\beta) - z^*(x_\beta))$ qui serviront à l'élaboration d'une variance expérimentale:

$$S_{kprior}(ivois) = \sum_{[V_\alpha(x_i)]_{étendu}} \frac{(z(x_\alpha) - z^*(x_\alpha))^2}{2Npoint} \qquad (b.26)$$

Suite à ces opérations effectuées sur un voisinage particulier, on itère celles-ci sur les autres zones tests.
Il en résulte autant de quantités $S_{kprior}(ivois)$ que de zones tests, ce qui permet d'établir leur moyenne:

$$E(k_{prior}) = \sum_{nvois} \frac{S_{kprior}(ivois)}{nvois} \qquad (b.27)$$

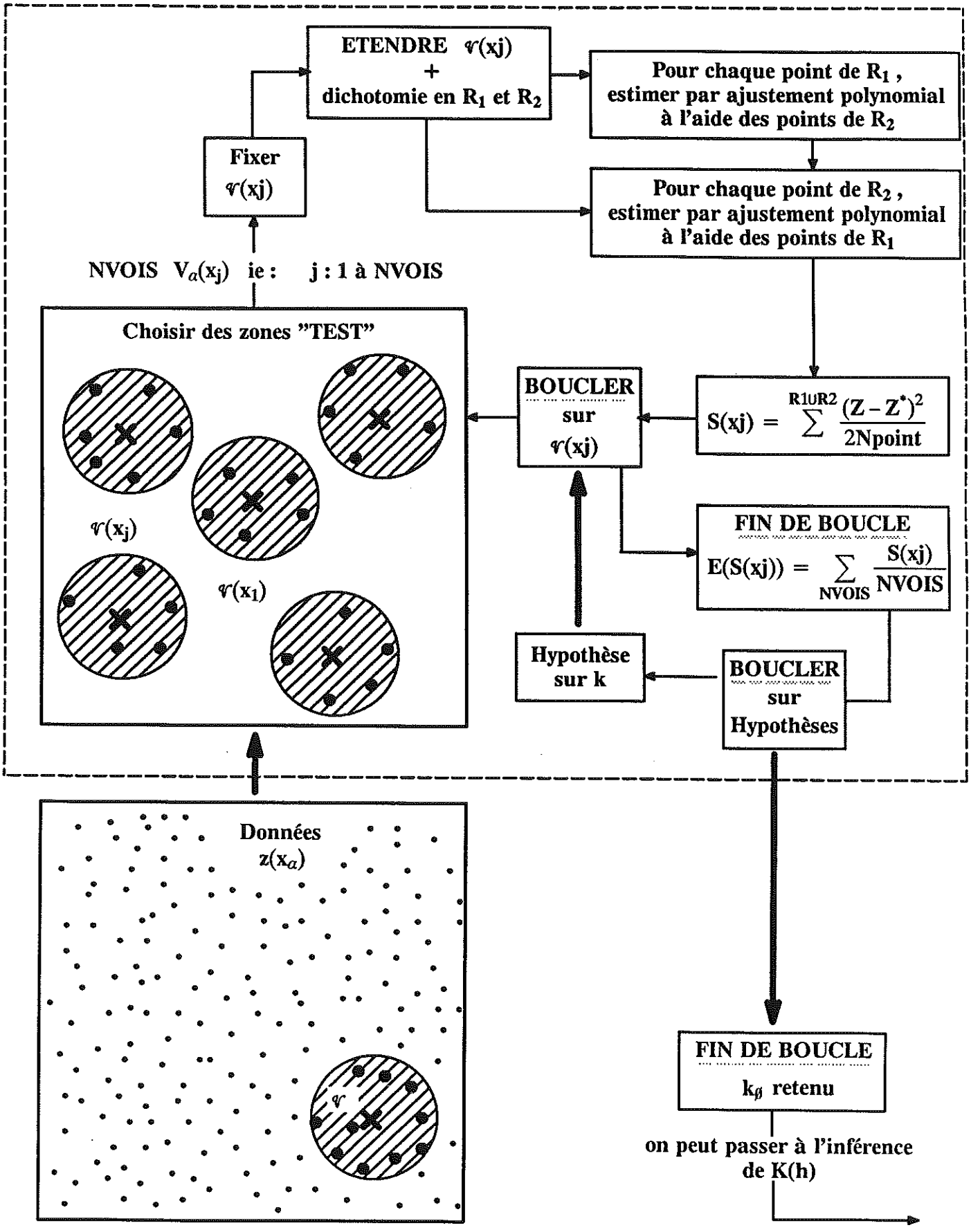


Figure b .2 : Inférence du degré polynomial (boucle-k)

C'est par b.27 que nous quantifions l'hypothèse "k = k_{prior}", et c'est en incrémentant k à partir de ∅ que nous obtenons le degré de la FAI-k: celui qui minimise b.27.

Si en théorie k peut varier de ∅ à l'infini, on se limite en pratique aux valeurs ∅ ,1 et 2. .

En réalité, ce ne sont pas véritablement les formules b.26 et b.27 qui sont utilisées, car elles s'avèreraient trop dépendantes des valeurs extrêmes prises par la variable. On leur substitue une anamorphose entière: celle qui affecte à l'erreur ($z(x_\emptyset)-z^*(x_\emptyset)$) la plus petite, l'étiquette 1, le rang 2 à la suivante, et ainsi de suite jusqu'à 2Npoint. C'est ensuite à partir de ces valeurs entières que les formules b.26 et b.27 sont calculées.

Inférence de K(h) (boucle-K) :

Une fois k déterminé pour une géométrie $\mathcal{V}$ fixée, il reste à déterminer le modèle de CG K(h) optimal pour ces caractéristiques $[\mathcal{V}\ ,\ k]$ (figure b.3).

Si l'ensemble des CG est vaste, on se limite en pratique à celles qui peuvent faire l'objet d'une inférence aisée. On choisit donc de se limiter au modèle spline:

$$K(h) = b|h|^2 \log|h| \qquad b \geq \emptyset \qquad \text{(b.17 , rappel)}$$

ainsi qu'aux modèles polynomiaux:

$$K(h) = -b_\emptyset pépite(h) + \sum_{p=\emptyset}^{k} (-1)^{p+1}\ b_{2p+1}\ |h|^{2p+1} \qquad b_p \geq \emptyset \qquad \text{(b.18 , rappel)}$$

formule dans laquelle la condition de positivité des coefficients est une condition suffisante mais non nécessaire pour assurer la positivité de K(h) (réf [**GST-8**]).

L'intérêt de ces modèles est que les inconnues b_p interviennent linéairement dans la formule.

La particularité de cet être mathématique K(h) est de ne pas modéliser la variabilité de Z(x), mais celle de CLA-k .

Pour l'inférer à partir des données dont nous disposons, il faudra construire des CLA-k $\sum_\alpha \lambda^\alpha Z(x_\alpha)$ telles que $\{\lambda_\alpha\}$ vérifie les contraintes b.14 , le degré k étant celui qui vient d'être déterminé. Or, on montre qu'un ajustement par "moindre-carré" permet de construire de telles CLA-k (réf [**GST-6**]). Il en résulte que la boucle-k met à notre disposition (2 x Nvois x Npoint) ensembles $\{\lambda_\alpha\}$ vérifiant b.14.

On fait donc une hypothèse sur K(h) en posant K(h) = $K_{prior}(h)$, avec pour $K_{prior}(h)$ une combinaison des modèles b.17 et b.18 , pour ensuite procéder comme précédemment par étape, le paramètre incrémental étant cette fois encore le voisinage:

- **Pour chaque voisinage,** on calcule la quantité $K_{prior}[\lambda_i]$:

$$K_{prior}[\lambda_i] = \quad K_{prior}[\{\lambda_\alpha\}_i] = \mathrm{Var}\left(\sum_\alpha \lambda^\alpha Z(x_\alpha)\right) = \sum_\alpha \sum_\beta \lambda^\alpha \lambda^\beta K^{\alpha\,\beta}_{prior} \qquad \text{(b.28)}$$

qui représente la variance théorique de la CLA-k associée à l'hypothèse K_{prior} et au voisinage "ivois". Cette expression contient des coefficients inconnus b_p .

- **Sur l'ensemble des voisinages,** on minimise la forme quadratique:

$$Q(b_p) = \sum_{j=1}^{2\ nvois} [\ S_k(ivois) - K_{prior}(\lambda_i)\ ]^2 \qquad \text{(b.29)}$$

expression dans laquelle S_k(ivois) représente la variance expérimentale b.26 associée à chaque $\mathcal{V}$. Cette minimisation permet d'inférer les coefficients b_p de $K_{prior}(h)$.

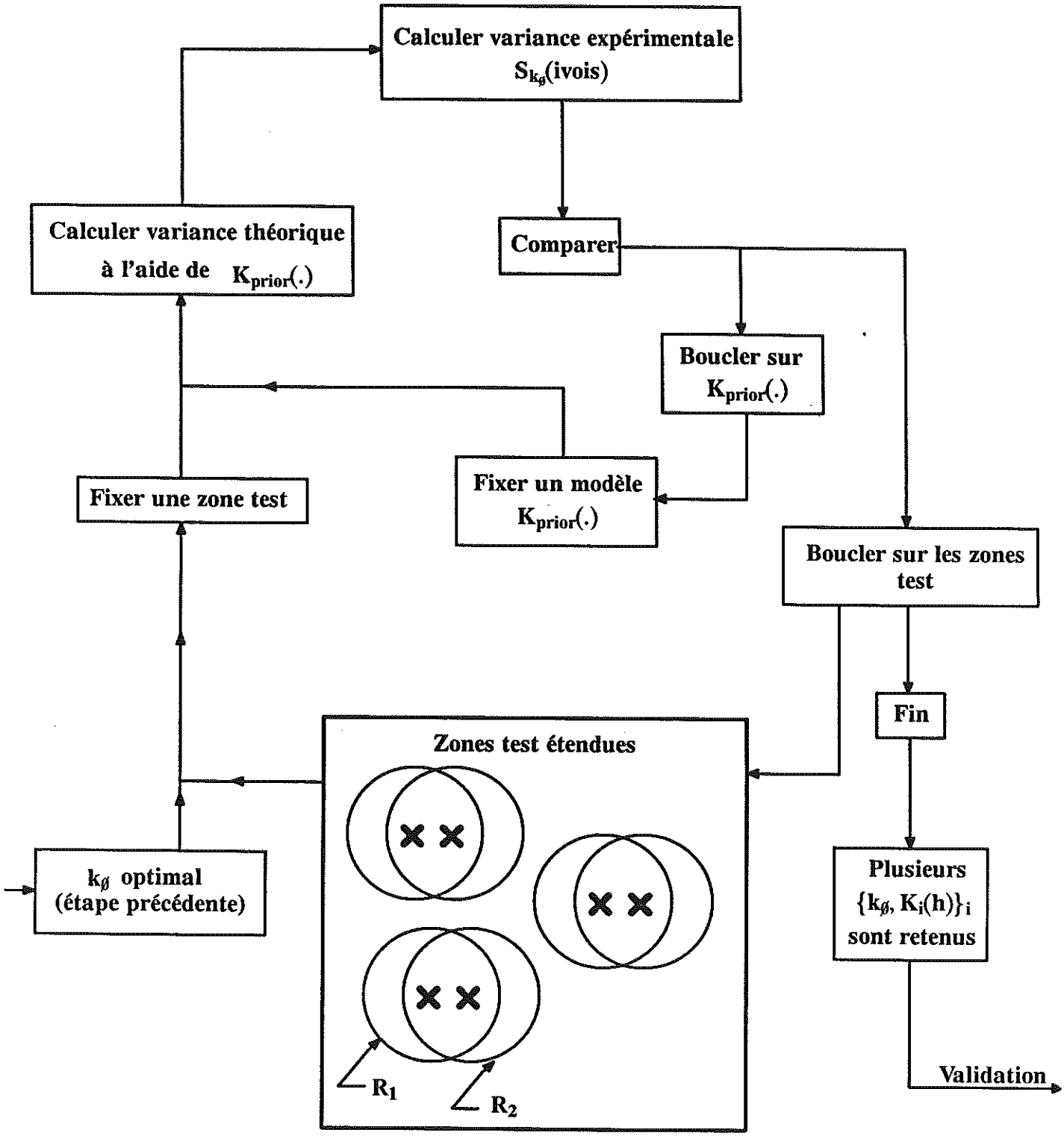


Figure b .3 : Inférence de la covariance généralisé K(h) (boucle-K)

Il reste à quantifier la qualité de l'hypothèse $K(h) = K_{prior}(h)$, et si la distance quadratique $Q(b_p)$ résultante de b.29 pourrait apparaître comme un bon test, on lui préfèrera le "Jack-knife" Q^*":

$$Q^* = 2r - \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2} \qquad (b-30)$$

avec:

$$r = \frac{\sum_{ivois} S_k(ivois)}{\sum_{ivois} K(\lambda_i)}$$: ratio entre variances expérimentales et théoriques

r_1 (respt. r_2) : ratio précédent établi sur $[\mathcal{V}(x_i)]_{R_1}$ (respt. $[\mathcal{V}(x_i)]_{R_2}$) seulement

n_1 (respt. n_2) : population de $[\mathcal{V}(x_i)]_{R_1}$ (respt. $[\mathcal{V}(x_i)]_{R_2}$)

Les modèles K(h) seront ensuite classés selon leurs Jack-knifes Q^* qui doivent être aussi proches de 1 que possible. Suite à quoi on peut se contenter du meilleur modèle en ce sens, ou le tester dans la "boucle-K^*" dont la description suit.

Validation du modèle (boucle-K^*)

Les étapes précédentes de l'inférence ont fourni des ensembles $\{\mathcal{V}, k, K_i(h)\}$ de paramètres. Pour déterminer l'ensemble optimal, on va mettre en oeuvre l'opérateur d'estimation caractérisé par cet ensemble de paramètres (figure b.4).

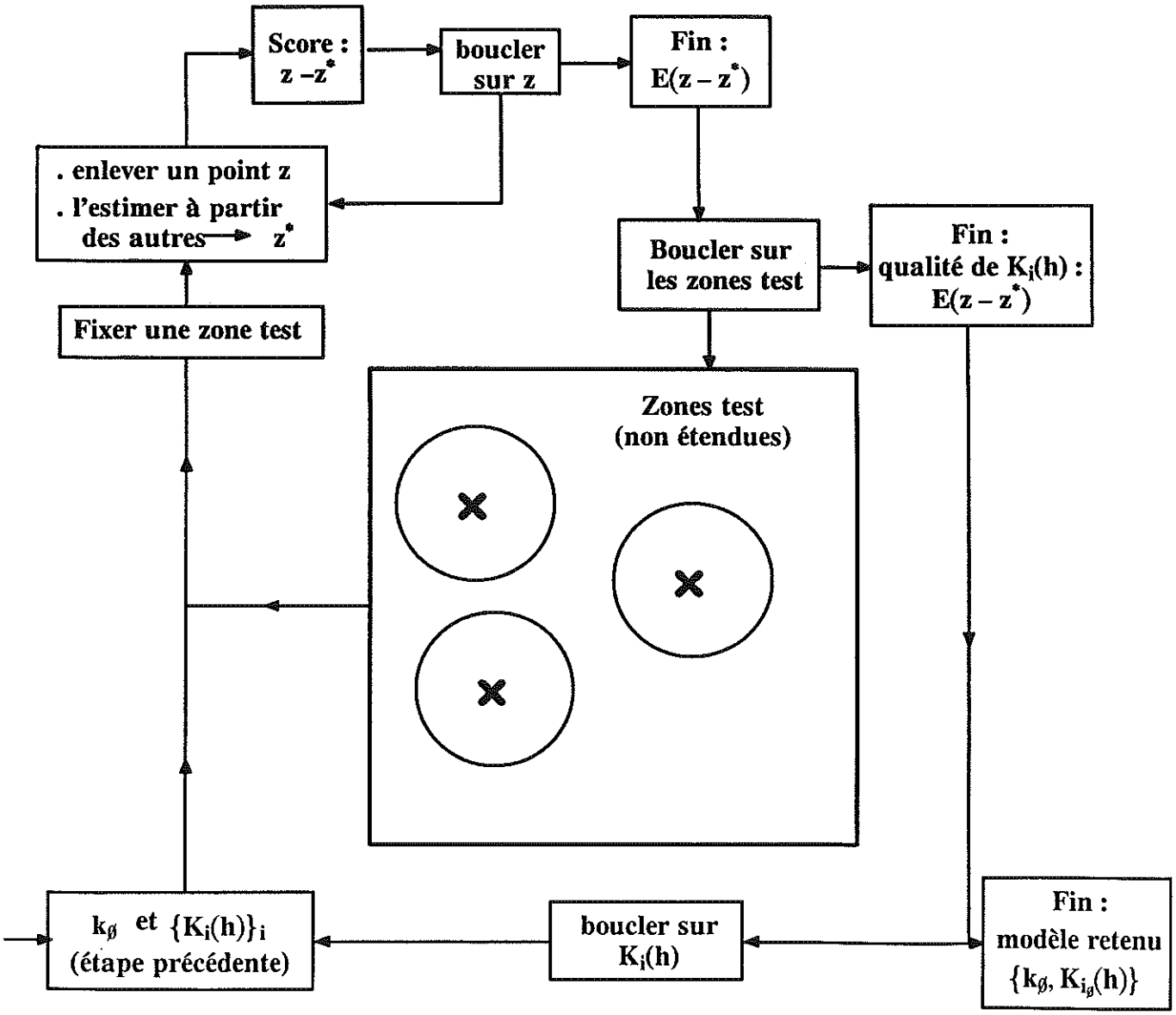


Figure b .4 : Validation du modèle (boucle-K^*)

Comme précédemment, nous opérerons par étapes successives dont le paramètre incrémental sera l'ensemble testé de paramètres. On commence donc par faire l'hypothèse que $\{\mathcal{V}\ ,\ k\ ,\ K(h)\} = \{\mathcal{V}\ ,\ k\ ,\ K_{i_\beta}(h)\}$, et on boucle sur les différentes zones tests $\mathcal{V}(x_i)$:

- **A $\mathcal{V}(x_i)$ fixé,** on applique les contraintes géométriques directement sur les (Npoint + 1) points les plus proches de la cible (qui est bien évidemment contenue dans l'ensemble), puis on supprime le point cible de l'ensemble, pour le réestimer.
  L'estimateur va fournir une valeur $z^*(x_\beta) = \text{Krig}(z(x_\beta))$ qui nous permettra d'établir un "score" (ie une erreur expérimentale) égal à $(z(x_\beta) - z^*(x_\beta))$. une erreur quadratique par $\mathcal{V}$ (en faisant varier $z(x_\beta)$ dans $\mathcal{V}$) et une erreur standardisée égale au ratio de l'erreur moyenne sur la variance théorique associée au modèle testé

- **En faisant varier $\mathcal{V}$** sur l'ensemble des zones tests, on est à même d'établir des moyennes de ces résultats.

Le modèle choisi est celui qui engendre la plus petite erreur quadratique moyenne.

Remarques :

Dans cette chaîne méthodologique qui a permis d'aboutir à l'ensemble $\{\mathcal{V}\ ,\ k\ ,\ K(h)\}$, les caractéristiques de $\mathcal{V}$ (diamètre, pas, population, forme géométrique...) n'ont pas fait l'objet d'une inférence. C'est pourquoi ce modèle n'est à ce stade qu'un ensemble $\{\mathcal{V}^{prior}\ ,\ k\ ,\ K(h)\}$. Mais rien n'empêche de réitérer l'ensemble de ces boucles, en faisant varier les caractéristiques de $\mathcal{V}$. Le seul problème est que cela s'avère d'une mise en oeuvre lourde.

C Krigeage (s)

Il n'existe pas "des" Krigeages mais "un" Krigeage puisque cette opération est avant tout un principe.

Dans ce qui suit, nous présentons le "principe Krigeage", puis nous montrons comment il est mis en oeuvre dans le cadre bien particulier des FAI-k-{f_i} (section C.2), du Krigeage de dérives (section C.3) et de l'Analyse Krigeante (section C.4).

C.1 Le principe "Krigeage"

Soit z(x,t) une variable informée d'une manière discrète et non nécessairement périodique dans l'espace-temps ("x" représente le vecteur de la localisation spatiale et "t" le temps de la mesure). Notre objectif est de former une estimation $z^*(x_i, t_i)$ de la valeur inconnue $z(x_i, t_i)$ en un point (x_i, t_i) quelconque du champ.

Choix d'un mode d'estimation :

● **La première étape** de la modélisation de ce problème d'estimation consiste (comme pour l'analyse) à interpréter z(x,t) comme la réalisation d'une Fonction Aléatoire (FA) Z(x,t).
Nous supposerons que la FA Z(.) possède une covariance non nécessairement stationnaire K(.)
On rappelle que cette covariance est définie par :

$$K(x_\alpha, t_\alpha \ ; \ x_\beta, t_\beta) = E\ [Z(x_\alpha, t_\alpha)\ Z(x_\beta, t_\beta)\ ] \qquad (c.1)$$

● **La deuxième étape** de la modélisation de ce problème d'estimation consiste à effectuer les <u>choix</u> suivants :

●● Elaborer $z^*(x_i, t_i)$ par une <u>combinaison linéaire</u> de la forme :

$$Z^*(x_i, t_i) = \sum_{\mathcal{V}} \lambda^\alpha z(x_\alpha, t_\alpha) \qquad (c.2)$$

avec :

$\mathcal{V} = \{\ z(x_\alpha, t_\alpha)\ \}$: voisinage de l'estimateur ie ensemble des points informants servant à l'estimation . ("Npoint" éléments). (c.3)

$\{\ \lambda^\alpha\ \}$: ensemble de pondérateurs inconnus. ("Npoint" éléments) (c.4)

●● Dire que dans le modèle, l'erreur aléatoire $[Z^*(x_i, t_i) - Z(x_i, t_i)\ ]$ possède une espérance nulle ie :

$$E\ [Z^*(x_i, t_i) - Z(x_i, t_i)\ ] = \emptyset \quad \forall\ (x_i, t_i) \qquad (c.5)$$

●● <u>Optimiser</u> l'estimateur en minimisant la variance de l'erreur $[Z^*(x_i, t_i) - Z(x_i, t_i)\ ]$.

Sous la contrainte c.5, cette variance possède alors pour expression :

$$VAR[Z^*(x_i, t_i) - Z(x_i, t_i)\ ] = \sum_{V_\alpha} \sum_{V_\beta} \lambda^\alpha\ \lambda^\beta\ K(x_\alpha, t_\alpha\ ;\ x_\beta, t_\beta)$$

$$- 2 \sum_{V_\alpha} \lambda^\alpha\ K(x_\alpha, t_\alpha\ ;\ x_i, t_i)$$

$$+ K(x_i, t_i\ ;\ x_i, t_i) \qquad (c.6)$$

Mise en oeuvre de ces principes :

Comme la covariance K(x,y) dépend de x et de y. on ne peut pas, à partir des données, l'inférer. De même, dire que c.5 est vraie peut s'avérer abusif et inadapté aux données. Pour mettre en oeuvre ces choix méthodologiques, il faut donc aller plus loin et formuler explicitement certaines hypothèses sur la FA Z(x,t). Ces hypothèses dépendent de la connaissance physique que nous avons du phénomène, ainsi que de ce que révèlent les données. Ainsi (par exemple), rien ne nous empêche de calculer le variogramme expérimental et de voir si celui-ci se prête à l'inférence d'un modèle stationnaire . Si tel est le cas, la covariance résultante K(x,y) ne dépend plus que de la seule distance |x-y|, ce qui permet de minimiser l'expression c.6 et d'aboutir à un système linéaire inversible.

En section C.2, nous élaborerons un système de Krigeage adapté au problème des perturbations diurnes en magnétisme.

Optimalité de ces choix méthodologiques :

En théorie, l'estimateur optimal de $z(x_i, t_i)$ à l'aide des données $\mathcal{V} = \{ z(x_\alpha, t_\alpha) \}$ est l'espérance conditionnelle :

$$z^*_{opt}(x_i, t_i) = E[Z(x_i, t_i) / \{\ldots, z(x_\alpha, t_\alpha), \ldots\}] \qquad \text{(c.8)}$$

L'estimateur $z^*_{opt}(x_i, t_i)$ donné par c.8 ne s'exprime comme une combinaison linéaire du type c.2 des points informants que si $\{ Z(x_i, t_i), \ldots\ldots, Z(x_\alpha, t_\alpha), \ldots. \}$ est multi-gaussien. Donc en théorie, notre choix de prendre un estimateur linéaire n'est optimal que dans un cadre gaussien. Nous justifions nos choix par les 2 points suivants :

- La mise en oeuvre de l'estimateur $z^*_{opt}(x_i, t_i)$ donné par c.8 nécessite l'inférence de la loi conditionnelle de $\{ Z(x_i, t_i), /\ldots\ldots, Z(x_\alpha, t_\alpha), \ldots. \}$.
  Cette inférence n'est pas possible dans la majorité des cas .
- On ne sait pas quantifier ce que coûte le choix d'un estimateur linéaire dans un cadre non gaussien.
  Ce choix est donc un pari .

C.2 Krigeage en FAI-k-{f_i}

Hypothèses sur la FA :

L'analyse des magnétismes spatiaux-temporels Krakatao Seaperc et Mohns (chap. 3) sont faites moyennant la dichotomie suivante de Z(x,t) :

$$Z(x, t) = Y(x) + D(t) \qquad \text{(c.9, rappel)}$$

avec :

$$D(t) = \sum_{i=\beta}^{Nfrèq} [A_i \cos(2\pi f_i t) + B_i \sin(2\pi f_i t)] \quad : \text{perturbations diurnes} \qquad \text{(c.10, rappel)}$$

$$Y(x) \quad : \text{anomalie spatiale, FAI-k de CG } K_Y(h) \qquad \text{(c.11, rappel)}$$

L'analyse de ce PTP fut alors conduite en considérant des CLA-k-{f_i} de la variable . Pour que l'erreur [$Z^*(x_j,t_j) - Z(x_j,t_j)$] soit elle-même une CLA-k-{f_i} , il est donc nécessaire que dans la combinaison linéaire $\sum_\alpha \lambda^\alpha z(x_\alpha, t_\alpha)$ qui constitue notre estimateur, les pondérateurs $\{ \lambda^\alpha \}$ vérifient :

$$\sum_\alpha \lambda^\alpha f_\ell(x_\alpha) = f_\ell(x_i) \qquad \forall\ \ell \le k \qquad \text{(c.12)}$$

$$\sum_{\alpha} \lambda^{\alpha} \cos(2\pi f_i t_{\alpha}) = \cos(2\pi f_i t_j) \qquad \forall\, i : 1 \ldots Nfrèq \qquad (c.13)$$

$$\sum_{\alpha} \lambda^{\alpha} \sin(2\pi f_i t_{\alpha}) = \sin(2\pi f_i t_j) \qquad \forall\, i : 1 \ldots Nfrèq \qquad (c.14)$$

Dans l'expression, c.12, on rappelle que les fonctions $f_\ell(x)$ désignent des polynômes spatiaux de degré inférieur à k.

Système de Krigeage en FAI-k-$\{f_i\}$:

Si $\{ \lambda^{\alpha} \}$ vérifie c.12, c.13 et c.14 , alors l'erreur d'estimation est une CLA-k-$\{f_i\}$ d'espérance nulle (dans le modèle), de variance ne s'exprimant qu'à l'aide de la CG K(h) et la minimisation sous contraintes de cette variance conduit au système suivant :

$$\sum_{\forall} \lambda^{\alpha} K_Y(h_{\alpha\beta}) + \sum_{l} \mu_l f_l(x_\beta) + \sum_{i} \mu_i^1 \cos(2\pi f_i t_\beta) + \sum_{i} \mu_i^2 \sin(2\pi f_i t_\beta)$$

$$= K_Y(h_{\beta j}) \;\; \forall\, \beta \qquad (c.15)$$

$$\sum_{\alpha} \lambda^{\alpha} f_\ell(x_\alpha) = f_\ell(x_i) \qquad \forall\, l \le k \qquad (c.12,\ rappel)$$

$$\sum_{\alpha} \lambda^{\alpha} \cos(2\pi f_i t_{\alpha}) = \cos(2\pi f_i t_j) \qquad \forall\, i : 1 \ldots Nfrèq \qquad (c.13,\ rappel)$$

$$\sum_{\alpha} \lambda^{\alpha} \sin(2\pi f_i t_{\alpha}) = \sin(2\pi f_i t_j) \qquad \forall\, i : 1 \ldots Nfrèq \qquad (c.14,\ rappel)$$

Ce système est appelé "système de Krigeage en FAI-k-$\{f_i\}$ " et noté dans la suite $\mathcal{S}\{ FAI-k-[f_i] \}$

Sa forme matricielle est :

$$\begin{pmatrix} K_Y(h_{\alpha\beta}) & 1 & \cos(2\pi f_i t_\beta) & \sin(2\pi f_i t_\beta) & f_\ell(x_\beta) \\ 1 & & & & \\ \cos(2\pi f_i t_\beta) & & & \emptyset & \\ \sin(2\pi f_i t_\beta) & & & & \\ f_\ell(x_\beta) & & & & \end{pmatrix} \begin{pmatrix} \lambda^{\alpha} \\ \mu_\emptyset \\ \mu_i^1 \\ \mu_i^2 \\ \mu_\ell \end{pmatrix} = \begin{pmatrix} K_Y(h_{j\beta}) \\ 1 \\ \cos(2\pi f_i t_j) \\ \sin(2\pi f_i t_j) \\ f_\ell(x_j) \end{pmatrix}$$

$K_Y(h_{\alpha\beta})$: *CG de la FAI-k Y(x)*

$\{ \lambda^{\alpha} \}$: *pondérateurs d'estimation*

$\{ \mu_\emptyset , .. \mu_i^1 , .. \mu_i^2 \}$: *pondérateurs de Lagrange*

$f_\ell(x_\beta)$: *polynômes spatiaux ($k < l+1$)*

$$Z(x, t) = Y(x) + S\mathcal{F}(t)$$

Estimer $z(x_j, t_j)$

Inversibilité et instabilités numériques :

Parce qu'elle est plongée dans l'espace-temps, la CG $K_Y(h)$ n'est plus que définie positive (non stricte). Cependant, l'inversibilité de $\mathcal{S}\{ FAI-k-[f_i] \}$ et l'unicité de la solution sont assurés ssi $K_Y(h)$ n'est pas périodique et si l'ensemble $\{ \ldots f_\ell(x_\alpha) \ ; \ \ldots \cos(2\pi f_i t_\alpha) \ ; \ \ldots \sin(2\pi f_i t_\alpha) \}$ est linéairement indépendant sur les données.

Cette indépendance linéaire signifie ici que si il existe un vecteur de coefficients $(\ldots C_l \ ; \ \ldots \ ; \ C_i^1 \ ; \ \ldots C_i^2)$ tels que pour tout $(x_\beta, t_\beta) \in [\alpha]$, :

$$\sum_\ell C_\ell f_\ell(x_\alpha) + \sum_i C_i^1 \cos(2\pi f_i t_\beta) + \sum_i C_i^2 \sin(2\pi f_i t_\beta) = \emptyset$$

alors : $(\ldots C_\ell \ ; \ \ldots \ ; \ C_i^1 \ ; \ \ldots C_i^2) = (\emptyset \ ; \ \ldots \emptyset \ ; \ \ldots \emptyset)$ (c.16)

Nous nous contenterons dans ce qui suit, d'examiner les configurations spatiales et temporelles qui détruisent l'indépendance linéaire. En ce qui concerne l'indépendance linéaire des polynômes du plan, ces configurations sont bien connues. Elles sont :

- En FAI-1, des données alignées sur une droite
- En FAI-2, des données alignées sur l'intersection d'une conique et d'un plan (2 droites parallèles, une ellipse, un cercle, etc ...)

En ce qui concerne l'indépendance linéaire des fonctions trigonométriques temporelles, ces configurations sont des prélèvements périodiques de D(t) à une cadence égale à la moitié d'une des périodes du système.

Si donc $\mathcal{V}$ possède une section spatiale ou temporelle qui répond exactement à l'une de ces géométries, le système n'est pas inversible.

Si les sections spatiale ou temporelle répondent approximativement à l'une de ces géométrie, la machine donnera une solution, mais celle-ci sera dénuée de sens et terriblement instable en ce sens qu'une légère modification sur la géométrie de $\mathcal{V}$ entraînera une grande modification de l'estimation. Ceci explique les problèmes rencontrés dans le chapitre 3 lors de l'analyse et de la modélisation du magnétisme spatio-temporel z(x,t) . En effet, les boucles d'inférence des caractéristiques de cette FA étaient faites en fonction d'un Krigeage en FAI-k-$\{f_i\}$. Or, dans ce problème, les échantillons étaient sur des profils sensiblement parallèles, de sorte que les géométries de type "spatial aveugle" (V_1^x) et dans une certaine mesure celles du type "spatial anisotrope " (V_2^x), étaient en moyenne approximativement des droites . Il en résultait que dans ces deux cas, l'inférence d'un degré $k > \emptyset$ était soit impossible, soit très instable numériquement, avec comme résultats des variances fortes et des conflits entre les divers tests. On comprend alors que la géométrie du type "spatial isotrope sautant" (V_3^x) soit celle qui ait donné les meilleurs résultats pour les 3 cas d'étude du chapitre 4, même si la géométrie temporelle résultante présentait parfois des dégénérescences qui portaient alors sur l'échantillonnage de la $S\mathcal{F}$ temporelle.

De même, on comprend maintenant les problèmes liés à la géométrie spatio-temporelle du type $V_4^{x\,t}$. En effet, si celle-ci assurait une indépendance linéaire dans le temps, elle engendrait dans l'espace une géométrie équivalente à une droite (celle des points contraints temporellement) qui traverse un amas de points contraints spatialement et de géométrie V_3^x. Si ces dernier points constituaient dans l'absolu une géométrie stable, elle cessait de l'être lorsqu'on leur adjoignait la ligne des points contraints temporellement. En effet cette superposition créait un contraste d'échelle qui réduisait (numériquement) la partie V_3^x de $V_4^{x\,t}$ à un seul point et de ce fait $V_4^{x\,t}$ à une droite. D'ou instabilité numérique.

C.3 Le Krigeage de dérives

Définition :

Des systèmes linéaires tels que $\mathcal{S}\{ FAI-k-[f_i] \}$ ont été obtenus par la minimisation d'une variance sous contrainte. Ces contraintes sont appelées "contraintes d'universalité" en Géostatistique et les fonctions analytiques (polynômes, séries trigonométriques, etc...) sur lesquelles elles portent sont appelées les "fonctions d'universalité" ou encore fonctions de dérive.

Les estimateurs obtenus par ces systèmes contiennent implicitement l'estimation de coefficients multiplicatifs des fonctions d'universalité. Il peut s'avérer utile d'examiner ces coefficients. Nous dirons que nous pratiquons le Krigeage d'une dérive lorsque nous examinons ces coefficients seuls ou multipliés par leurs fonctions d'universalité .

Krigeage Trigonométrique :

Examinons le système $\mathcal{S}\{\ FAI-k-[f_i]\ \}$. En changeant le second membre de la manière suivante :

$$\begin{pmatrix} \vdots \\ K_Y(h_j\ \beta) \\ \hline 1 \\ \hline \cos(2\pi f_i t_j) \\ \hline \sin(2\pi f_i t_j) \\ \hline f_\ell(x_j) \end{pmatrix} \rightarrow \begin{pmatrix} \vdots \\ \emptyset \\ \hline 1 \\ \hline \cos(2\pi f_i t_j) \\ \hline \sin(2\pi f_i t_j) \\ \hline \emptyset \end{pmatrix} \qquad \text{(c.17)}$$

et en conduisant la résolution du système, on obtient une estimation de la S $\mathcal{F}$ associée à D(t) dans la dichotomie c.9. Lorsque le système $\mathcal{S}\{\ FAI-k-[f_i]\ \}$ est utilisé avec le second membre c.17, nous disons que nous effectuons un Krigeage Trigonométrique (noté KT) .

Précautions d'emploi :

Le Krigeage de dérives est à manier avec précautions. Il ne possède de sens que s'il existe véritablement, dans la FA qui est analysée, une composante qui possède un comportement comparable à celui des fonctions universelles qui lui sont associées.

Ainsi, si Z(x,t) est le magnétisme spatio-temporel, on sait que la S $\mathcal{F}$ dans la dichotomie c.9 donne une représentation acceptable du comportement temporel de cette VR si les fréquences utilisées sont 24^h, 12^h, etc ..

De ce fait, le système $\mathcal{S}\{\ FAI-k-[f_i]\ \}$ avec le second membre c.17 constitue un estimateur acceptable des perturbations diurnes D(t) à partir du magnétisme spatio-temporel z(x,t) . C'est ce système qui est utilisé dans le chapitre 4 pour effectuer un filtrage sur Krakatao, Seaperc et Mohns .

Par contre, il existe des cas où le krigeage de dérive est inopportun. Réexaminons le système $\mathcal{S}\{\ FAI-k-[f_i]\ \}$ et effectuons le krigeage d'une dérive spatiale . Pour ce faire, modifions le second membre du système comme suit :

$$\begin{pmatrix} \vdots \\ K_Y(h_j\ \beta) \\ \hline 1 \\ \hline \cos(2\pi f_i t_j) \\ \hline \sin(2\pi f_i t_j) \\ \hline f_\ell(x_j) \end{pmatrix} \rightarrow \begin{pmatrix} \vdots \\ \emptyset \\ \hline 1 \\ \hline \emptyset \\ \hline \emptyset \\ \hline f_\ell(x_j) \end{pmatrix} \qquad \text{(c.18)}$$

En Théorie, l'estimation obtenue ne correspond à rien du phénomène puisque par définition, une FAI-k est déterminée à un polynôme de degré k près. S'il est possible qu'à l'échelle de $\mathcal{V}$, la VR présente occasionnellement un comportement polynomial, ceci a peu de chances de se produire partout dans le champ, et le krigeage de la dérive ne risque de refléter en sortie du système ... que les fonctions arbitraires mises en entrée ...

Ce mode de krigeage nécessite donc une bonne connaissance de la physique du phénomène pour pouvoir juger de l'adéquation d'une fonction universelle à une "tendance réelle".

C.4 L'analyse krigeante

Définition :

Lors de l'analyse des séries temporelles Kakioka et Huancayo, le variogramme de la FA Z(.) présentait la forme additive suivante :

$$\gamma_Z(t) = \sigma_Y(t) + \gamma_{\text{pério}}(t) \qquad \text{(c.19)}$$

avec :

$$\gamma_{\text{pério}}(t) = \sum_n \xi_n \ (1 - \cos(2\pi f_n t)) \qquad \xi_n \geq \emptyset \qquad \text{(c.20)}$$

D'une manière plus générale, un ajustement de variogramme peut conduire à une somme de structures (deux dans l'exemple, mais non limitatif bien entendu) :

$$\gamma_Z(.) = \gamma_{Y_1}(.) + \gamma_{Y_2}(.) \qquad \text{(c.21)}$$

et il est bien entendu possible d'associer à cette sommation son équivalent en terme de variables statistiquement orthogonales :

$$Z(.) = Y_1(.) + Y_2(.) \qquad \text{(c.22)}$$

avec :

$Y_1(.)$: FA de variogramme $\gamma_{Y_1}(.)$ (c.23)

$Y_2(.)$: FA de variogramme $\gamma_{Y_2}(.)$ (c.24)

Il peut s'avérer utile d'estimer $Y_1(.)$ ou $Y_2(.)$. Nous disons dans ce cas que nous pratiquons une Analyse Krigeante (notée "AK")

Intérêt de L'A.K. :

Le problème ici est de savoir si la dichotomie c.22 possède un sens physique.

Dans le cas de c.19, on sait que $\gamma_{\text{perio}}(.)$ représente dans une série temporelle de perturbations, la variabilité des fréquences qui "émergent" du spectre (ie celles à fortes énergies).

De même, lorsque Z(.) représente le magnétisme spatial et que $\gamma_{Y_1}(.)$ et $\gamma_{Y_2}(.)$ sont des fonctions à portées a_{Y_i} telles que (par exemple) $a_{Y_1} \ll a_{Y_2}$, on sait qu'alors cette réalité expérimentale peut s'interpréter comme la présence dans Z(.) d'anomalies locales $Y_1(.)$ et d'anomalies régionales $Y_2(.)$ (ref [**GST-3**]).

Systèmes d'Analyse Krigeante Trigonométrique :

Comme pour le krigeage de dérives, la mise en oeuvre de l'AK se pratique par une modification des seconds membres des systèmes .

Mais alors que cette modification porte pour le krigeage de dérives sur la partie universelle, elle porte pour l'AK sur la partie covariance de ce second membre. Ainsi, reprenons l'exemple d'une

série temporelle de perturbation (formules c.19 et c.20). Le système qui lui est associé pour estimer $z(t_i)$ dans ce cadre temporel est :

$$\left(\begin{array}{c|c} \begin{array}{c} \gamma_Z(t_\alpha\,\beta) \\ = \quad \sigma_Y(t_\alpha\,\beta) \\ + \\ \sum_n \xi_n(1-\cos(2\pi f_n t_\alpha\,\beta)) \end{array} & \begin{array}{c} \vdots \\ 1 \\ \vdots \end{array} \\ \hline \text{----- 1 -----} & \emptyset \end{array}\right)\left(\begin{array}{c} \vdots \\ \lambda^\alpha \\ \vdots \\ \hline \mu_\beta \end{array}\right) = \left(\begin{array}{c} \gamma_Z(t_i\,\beta) \\ = \quad \sigma_Y(t_i\,\beta) \\ + \\ \sum_n \xi_n(1-\cos(2\pi f_n t_i\,\beta)) \\ \vdots \\ \hline 1 \end{array}\right)$$

Il est immédiat de voir qu'en changeant le second membre du système comme suit :

$$\left(\begin{array}{c} \gamma_Z(t_i\,\beta) \\ = \quad \sigma_Y(t_i\,\beta) \\ + \\ \sum_n \xi_n(1-\cos(2\pi f_n t_i\,\beta)) \\ \vdots \\ \hline 1 \end{array}\right) \rightarrow \left(\begin{array}{c} \sum_n \xi_n(1-\cos(2\pi f_n t_i\,\beta)) \\ \vdots \\ \hline 1 \end{array}\right) \qquad \text{(c.25)}$$

on obtient une estimation de la composante périodique de $Z(t_i)$.

Un autre cas intéressant est celui où la $S\mathcal{F}$ associée à un variogramme expérimental n'engendre que des énergies négatives. Dans ce cas, le système d'estimation du phénomène peut s'exprimer uniquement à l'aide de séries de fourier (celles issues de la TFD du variogramme) :

$$\left(\begin{array}{c|c} \begin{array}{c} \text{-----} \\ \sum_n \xi_n(1-\cos<2\pi f_n,\, h_\alpha\,\beta>) \\ \text{-----} \end{array} & \begin{array}{c} \vdots \\ 1 \\ \vdots \end{array} \\ \hline \text{----- 1 -----} & \emptyset \end{array}\right)\left(\begin{array}{c} \vdots \\ \lambda^\alpha \\ \vdots \\ \hline \mu_\beta \end{array}\right) = \left(\begin{array}{c} \vdots \\ \sum_n \xi_n(1-\cos<2\pi f_n,\, h_i\,\beta>) \\ \vdots \\ \hline 1 \end{array}\right)$$

Moyennant le changement du second membre du système comme suit :

$$\left(\begin{array}{c} \sum_n \xi_n(1-\cos<2\pi f_n,\, h_i\,\beta>) \\ \vdots \\ \hline 1 \end{array}\right) \rightarrow \left(\begin{array}{c} \xi_\beta(1-\cos<2\pi f_\beta,\, h_i\,\beta>) \\ \vdots \\ \hline 1 \end{array}\right) \qquad \text{(c.26)}$$

on obtient une estimation de l'atome particulier associée à la fréquence spatiale f_β .

Si l'intérêt de ce dernier système n'est pas immédiat dans une estimation à voisinage unique (du fait que l'amplitude se déduit de l'énergie ξ_β), on voit dans le chapitre 5 qu'en voisinage glissant, il permet notamment d'estimer les zones de transition d'une image (ie les hautes fréquences spatiales) si la cible f_β est élevée.

Parce que les seconds membres c.25 et c.26 ne contiennent que des fonctions de variogramme périodiques , les systèmes d'AK associés sont appelés systèmes d'"Analyse Krigeante Trigonométrique" (noté AKT) .

D Détails de la modélisation de la campagne Krakatao

Ce qui suit donne le détail de l'inférence des modèles qui ont permis le filtrage des perturbations dans le chapitre 4. Une grande attention est portée à l'inférence des paramètres.

D.1 Séquencement en "pseudo-profils"

Pour mener à bien l'analyse et la modélisation de ce PTP, nous adopterons 2 types de séquencement de la route du bateau :

- un premier qui s'identifie à de véritables profils, et dont la longueur (Δcode) sera de 12 heures. $\mathcal{V}$ sera alors du type $\mathcal{V}_2^x$
- un deuxième dont la longueur Δcode est de 1 heure, et qui conduira à des géométries de type $\mathcal{V}_3^x$ et $\mathcal{V}_4^{x\,t}$

D.2 Inférence de $\mathcal{V}$, "k" , $\{f_i\}$

Nous allons ici conditionner l'inférence des paramètres { k , $\{f_i\}$ } à certaines géométries particulières de $\mathcal{V}$. Les configurations $\mathcal{V}$ possibles seront prises dans l'ensemble { $\mathcal{V}_2^x$, $\mathcal{V}_3^x$, $\mathcal{V}_4^{x\,t}$ } (chapitre 3). C'est pour chacun de ces éléments qu'on cherchera à trouver le degré "k" (éventuellement nul) des polynômes de l'espace, et le sous-ensemble $\{f_i\}$ (éventuellement vide) des fréquences temporelles qui permettent au mieux d'ajuster la tendance de z(x, t) à l'échelle de $\mathcal{V}$.

Ces opérations s'effectuent en diverses zones test du champ. Rappelons à ce propos que l'extension spatiale qui est faite sur chaque voisinage test et la dichotomie en deux sous-ensembles peuvent très bien détruire la géométrie initiale fixée.

Rappelons également que pour chaque hypothèse $\{\mathcal{V}, k, \{fi\}\}_{prior}$ testée, deux critères sont évalués :

- La distance quadratique $E\,(z - z^*)^2$ (notée critère "B" dans la suite) établie sur les "Nvois" zones test du champ.
  Dans ce critère, on rappelle que z^* est l'estimation obtenue par moindre carré à l'aide des fonctions de dérive de l'hypothèse, et z la valeur connue, mais retirée de $\mathcal{V}$ le temps d'une estimation.
  Ce critère est a priori très dépendant des variances anormalement fortes qui peuvent survenir, mais il donne un ordre de grandeur des variances en jeu.
- Le rang moyen (noté critère "A" dans la suite) obtenu par les "Nvois" zones test du champ.
  Pour chaque ajustement local, les différentes hypothèses sont étiquetées par un entier compris entre 1 et le nombre d'hypothèses.
  Ainsi, une hypothèse qui obtient 1 sera celle ayant engendré la distance quadratique minimale comparativement aux autres hypothèses.
  Ce critère est dit "stable" car peu dépendant de certains comportements aberrants de la variable (ou de son estimateur) .

Bien qu'apparemment redondant, l'examen conjoint des critères A et B peut s'avérer utile pour analyser le détail de ces statistiques. C'est cet examen qui permettra de choisir l'ensemble { $\mathcal{V}$,k, {fi}} .

Inférence de { k , {fi} } conditionnellement à $\mathcal{V}_2^x$:

La décomposition de la route est celle des "pseudo-profils" lâches (Δcode = 12^h). La population de $\mathcal{V}$ a été fixée à Npoint = 20 éléments dont 7 qui doivent appartenir à 7 profils différents (Ncode = 7). La configuration résultant de ces contraintes appliquées aux données sur profils est donc celle du type $\mathcal{V}_2^x$ (spatial-anisotrope).

Conditionnellement à cette géométrie, les 6 ensembles suivants de paramètres ont été testés :

- k = ∅ , Z(x,t) FAI-∅ , spatiale.
  La seule fonction constante ("1") modélise la dérive spatiale, et aucune fonction de dérive temporelle n'est utilisée.

- k = 1 , Z(x,t) FAI-1 , spatiale.
  Les fonctions 1, x et y modélisent la dérive spatiale et aucune fonction de dérive temporelle n'est utilisée.
- k = 2 , Z(x,t) FAI-2 , spatiale.
  Les fonctions 1, x, y, x^2, y^2 et xy modélisent la dérive spatiale et aucune fonction de dérive temporelle n'est utilisée.
- k = ∅ ou 1 ou 2 et {fi} = {24^h}, Z(x,t) FAI-k-{24h} , spatio-temporelle.
  La dérive spatiale est modélisée comme précédemment, mais avec chaque fois les fonctions cosinus et sinus de la seule période 24^h comme fonctions de dérive temporelle.

Ces 6 ensembles de paramètres ont été testés sur 180 voisinages répartis d'une manière homogène sur la zone d'étude. Il en résulte 360 tests statistiques .

Si 360 voisinages sont ainsi testés, pour chacun d'eux, 20 ajustements (variable npoint) sont faits, de sorte que les statistiques sont basées sur 7 200 écarts ($z - z^*$).

(A) Classement par rang minimal (test stable)

Rang moyen	Fonctions de dérive	
	spatiales	temporelles
3.26	*1,x,y,x^2,y^2,xy*	
3.39	*1,x,y,x^2,y^2,xy*	*24^h*
3.82	*1,x,y*	*24^h*
3.98	*1,x,y*	
4.21	*1*	*24^h*
4.34	*1*	

(B) Classement par erreur quadratique minimale (test dépendant des hétérogénéités)

$E(Z - Z_*)^2$ nT^2	Fonctions de dérive	
	spatiales	temporelles
139	*1*	*24^h*
145	*1,x,y,x^2,y^2,xy*	*24^h*
157	*1,x,y*	
192	*1*	
192	*1,x,y*	*24^h*
767	*1,x,y,x^2,y^2,xy*	

meilleur compromis ——> adopté

conflit ——> refusé

Figure d.1 : KRAKATAO – Magnétisme : Détection de fréquences temporelles (voisinage du type $\mathcal{V}_2^x$)

L'ensemble des résultats figure en d.1.
Si on se contente d'examiner le critère par rang minimal, on constate que parmi ces 6 ensembles, l'ensemble optimal est celui pour lequel k = 2 et qui ne contient pas de dérive temporelle.

On notera que son rang moyen n'est (approximativement) que de 3 , pour un intervalle de variations de [3.26, 4.34], ce qui est faible. Ceci signifie que cet ensemble de fonctions est le plus stable, mais il est peu précis. Cette dernière remarque est confirmée par l'examen du critère quadratique, où cette fois-ci, ce même ensemble se retrouve en dernière position avec une variance expérimentale $E(z - z^*)^2$ de 767 nT^2. Cette valeur très élevée par rapport aux autres, confirme le manque de précision de l'ensemble {1, x, y, x^2, y^2, xy}. Il y a donc conflit entre ces deux critères, puisque ce qui est optimal pour l'un ne l'est pas pour pour l'autre. L'explication proposée est la suivante : si les fonctions spatiales sont peu dépendantes de certains comportements aberrants de la variable, elles s'avèrent peu précises car elles ne parviennent pas à expliquer la tendance temporelle du phénomène. Il est très probable qu'ici, ces

comportements "aberrants" soient tout simplement les écarts aux points de recoupement des profils transverses et des profils longitudinaux. Dans ces zones particulières, l'ajustement par moindre carré donne alors un résultat qui n'est qu'un compromis peu précis entre des données fortement hétérogènes. Ce compromis, parce qu'il est peu précis, engendre alors une forte variance expérimentale.
L'introduction de $T = 24^h$ dans le système d'analyse permet d'expliquer ces comportements et donc de diminuer d'autant les variances expérimentales puisque les ensembles $\{ 1, T=24^h \}$ et $\{ 1, x, y, x^2, y^2, xy$ et $T=24^h \}$ permettent d'obtenir des variances respectivement égales à $139\ nT^2$ et $145\ nT^2$, soit près d'un sixième seulement de la variance précédente.
Comme corrélativement ce gain de précision pour ces deux derniers ensembles s'accompagne d'une dégradation des rangs moyens, ceci signifie que ce système d'analyse est très instable. Il est fort probable que ces instabilités soient dues à la configuration V_2^x dont on sait qu'elle échantillonne mal la tendance temporelle. Il en résulte que si ces deux derniers ensembles offrent bien souvent un gain de précision, ces instabilités numériques leur donnent cependant des rangs moyens non minimaux. Mais comme malgré tout les variances sont correctes, cela signifie que l'augmentation de variance issue de ces instabilités est moindre comparée à l'augmentation de variance issue de l'ensemble $\{1, x, y, x^2, y^2, xy\}$ de paramètres.

Compte tenu de cette analyse, et du faible domaine de variation des rangs, nous avons choisi d'adopter un compromis entre ces deux critères, et de prendre l'ensemble: $\{ 1, x, y, x^2, y^2, xy$ et $T=24^h \}$ (ie $Z(x,t)$ FAI-2-$\{24^h\}$, spatio-temporelle)., qui représentera l'optimalité conditionnellement à la géométrie du type V_2^x.

Inférence de $\{ k, \{fi\} \}$ conditionnellement à $\mathcal{V}_3^x$:

La décomposition de la route est celle de pseudo-profils très courts (Δcode = 1 heure). La population de $\mathcal{V}$ a été fixée à Npoint=30 éléments qui doivent appartenir à 30 pseudo profils différents (Ncode = 30). La configuration résultante est donc celle du type $\mathcal{V}_3^x$ (spatial-isotrope-sautant).

Conditionnellement à cette géométrie, les 9 ensembles suivants de paramètres ont été testés :

- $\{ f_L = 1, T_1 = 24^h \}$, $Z(x,t)$ FAI-$\emptyset$-$\{24^h\}$, spatio-temporelle
- $\{ f_L = 1 \}$, $Z(x,t)$ FAI-$\emptyset$, spatiale
- $\{ f_L = 1, T_1 = 24^h$ et $T_2 = 12^h \}$, $Z(x,t)$ FAI-$\emptyset$-$\{24^h,12^h\}$, spatio-temporelle
- $\{ f_L = 1, x, y \}$, $Z(x,t)$ FAI-1$\}$, spatiale
- $\{ f_L = 1, x, y\ T_1 = 24^h \}$, $Z(x,t)$ FAI-1-$\{24^h\}$, spatio-temporelle
- $\{ f_L = 1, x, y, T_1 = 24^h$ et $T_2 = 12^h \}$, $Z(x,t)$ FAI-1-$\{24^h, 12^h\}$, spatio-temporelle
- $\{ f_L = 1, x, y, x^2, y^2, xy \}$, $Z(x,t)$ FAI-2$\}$, spatiale
- $\{ f_L = 1, x, y, x^2, y^2, xy, T_1 = 24^h \}$, $Z(x,t)$ FAI-2-$\{24^h\}$, spatio-temporelle
- $\{ f_L = 1, x, y, x^2, y^2, xy, T_1 = 24^h$ et $T_2 = 12^h\}$, $Z(x,t)$ FAI-2-$\{24^h,12^h\}$, spatio-temporelle

Comme précédemment, ces 9 ensembles de paramètres ont été testés sur 180 voisinages répartis d'une manière homogène sur la zone d'étude. L'ensemble des résultats figure en d.2 .

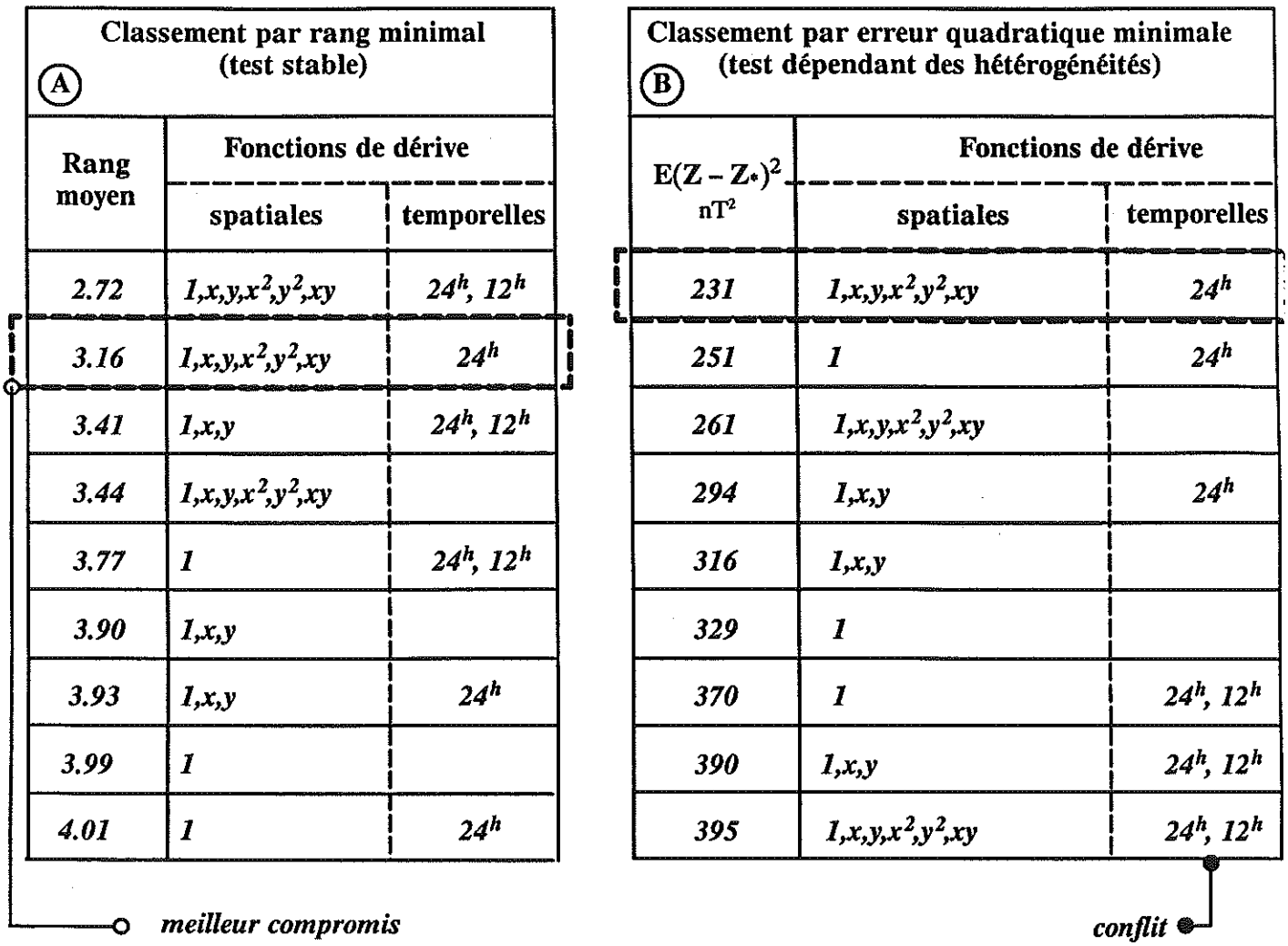


Figure d.2 : Détection de fréquences temporelles (voisinage du type $\mathcal{V}_3^x$)

On se retrouve dans une position conflictuelle comparable au cas précédent, puisque l'ensemble $\{1, x, y, x^2, y^2, xy, T_1 = 24^h$ et $T_2 = 12^h\}$ est optimal pour le test stable, avec un rang moyen de 2.72 , mais donne les plus mauvais résultats en terme de variance expérimentale de (395 nT^2). On notera cependant que l'intervalle de variations de la distance quadratique est ici [231 nT^2, 395 nT^2] alors qu'il était de [139nT^2 , 767 nT^2] précédemment.

On notera également que pour des ensembles identiques (par exemple l'ensemble $\{1, x, y, x^2, y^2, xy, T_1 = 24^h\}$) mais pour 2 géométries différentes, les distances quadratiques obtenues sont différentes, selon que $\mathcal{V}$ est du type $\mathcal{V}_2^x$, (on a alors $E(z - z^*)^2 = 145\ nT^2$) ou du type $\mathcal{V}_3^x$ (il vient alors $E(z - z^*)^2 = 231\ nT^2$).

Si on s'interroge sur l'opportunité d'introduire ou non $T_2 = 12^h$ dans le système d'analyse, la réponse est apportée, non par le critère de rang qui montre peu de différences (à f_L fixé), mais bien par le critère des distances quadratiques où les scores sont nettement favorables à $T_1 = 24^h$ seule , reléguant $T_2 = 12^h$ dans la queue des scores. On peut raisonnablement en déduire que la période 12^h n'a pas été détectée. On renoncera donc à cette dernière période.

Aux mêmes maux les mêmes remèdes, à savoir l'adoption de l'ensemble suivant :

$\{ 1, x, y, x^2, y^2, xy$ et $T=24^h \}$ (ie Z(x,t) FAI-2-$\{24^h\}$, spatio-temporelle).,

qui représentera l'optimalité conditionnellement à une géométrie du type $\mathcal{V}_3^x$. Si cet ensemble est identique au précédent, on notera qu'il se trouve ici en première position pour le critère de variance expérimentale (alors qu'il n'était qu'en seconde position précédemment).

Inférence de { k , {fi} } conditionnellement à $\mathcal{V}_4^{x\,t}$:

La décomposition en pseudo-profils est la même que précédemment (Δcode = 1 heure). La population de $\mathcal{V}$ est fixée à 30 points, dont 20 seulement doivent être répartis sur 20 "pseudo" profils différents. Les 10 points complémentaires seront contraints à encadrer (temporellement) t_x (temps ciblé par l'analyse) à maille régulière centrée sur t_x et de pas = 1 h. La configuration résultante est donc celle du type $\mathcal{V}_4^{x\,t}$ (spatio-temporel).

On rappelle que si la partie spatiale de $\mathcal{V}_4^{x\,t}$ est sautante, son complément temporel est glissant avec t_x. Conditionnellement à cette géométrie les 9 ensembles de paramètres à tester sont ceux de l'analyse précédente. L'ensemble des résultats figure en d.3.

Classement par rang minimal (test stable) Ⓐ

Rang moyen	Fonctions de dérive	
	spatiales	temporelles
2.69	*1,x,y,x^2,y^2,xy*	*24^h, 12^h*
3.04	*1,x,y*	*24^h, 12^h*
3.37	*1,x,y,x^2,y^2,xy*	*24^h*
3.45	*1,x,y,x^2,y^2,xy*	
3.71	*1*	*24^h, 12^h*
3.84	*1,x,y*	*24^h*
3.84	*1,x,y*	
4.31	*1*	
4.36	*1*	*24^h*

Classement par erreur quadratique minimale (test dépendant des hétérogénéités) Ⓑ

$E(Z-Z_*)^2$ / nT^2	Fonctions de dérive	
	spatiales	temporelles
224	*1*	*24^h*
270	*1,x,y*	*24^h*
304	*1*	*24^h, 12^h*
316	*1,x,y*	
324	*1,x,y*	*24^h, 12^h*
335	*1*	
355	*1,x,y,x^2,y^2,xy*	
364	*1,x,y,x^2,y^2,xy*	*24^h, 12^h*
660	*1,x,y,x^2,y^2,xy*	*24^h, 12^h*

Figure d.3 : Détection de fréquences temporelles (voisinage du type $\mathcal{V}_4^{x\,t}$)

Une première remarque qui s'impose est liée au nombre de voisinages qui ont pu être construits en respectant ces contraintes très strictes : 35 seulement (soit du fait de l'extension, 70 tests), Dès lors, les résultats qui suivent ne sont pas véritablement statistiques, puisque établis sur un effectif trop réduit.

Si 35 voisinages seulement ont pu être construits, c'est parce qu'on impose à la machine de respecter très exactement les contraintes temporelles du voisinage, ce qui n'est pas toujours possible.

L'examen de la figure précitée révèle que nous sommes dans une configuration extrême puisque :

- L'ensemble { 1, x, y, x^2, y^2, xy, 24^h, 12^h } est optimal par son rang (valeur 2.69), mais il est désastreux en terme de variances expérimentales (valeur 660 nT^2)
- L'ensemble {1, 24^h} est optimal par sa variance expérimentale (valeur 224 nT^2), mais il est désastreux en terme de rang moyen (valeur 4.36).
- Pour tout autre ensemble de paramètres, le conflit subsiste.

Il en résulte que l'adoption d'un compromis ne signifie rien ici, puisque tout ensemble est équivalent en terme de juste équilibre entre rang moyen et variance expérimentale. Cette situation signifie qu'il y a une totale incompatibilité entre les tests à notre disposition et les modèles proposés. On se doit donc d'ores et déjà de renoncer à $\mathcal{V}_4^{x\,t}$. C'est donc uniquement par curiosité que nous poursuivrons l'analyse conditionnelle à cette géométrie. Nous adopterons pour cela l'ensemble :

{ 1, x, y, x^2, y^2, xy, $T_1 = 24^h$, $T_2 = 12^h$ }$_{prior}$, Z(x,t) FAI-2-{24^h,12^h}, spatio-temporelle,

qui ne représente aucune optimalité, qu'on associera cependant à un voisinage de type $\mathcal{V}_4^{x\,t}$ et dont on sait d'ores et déjà que les résultats ne seront pas bons.

L'indice "prior" est là pour rappeler que ce modèle n'a pas été inféré .

Bilan :

Afin de boucler cette analyse, il convient de comparer entre eux les trois ensembles de paramètres qui s'en dégagent:

- { $\mathcal{V} = \mathcal{V}_2^x$; 1, x, y, x^2, y^2, xy ; T= 24^h }
- { $\mathcal{V} = \mathcal{V}_3^x$; 1, x, y, x^2, y^2, xy ; T= 24^h }
- { $\mathcal{V} = \mathcal{V}_4^{x\,t}$; 1, x, y, x^2, y^2, xy ; T= 24^h, 12^h}$_{prior}$

Pour mener à bien cette comparaison, seul le critère de variance expérimentale est exploitable puisque celui des rangs est relatif au nombre d'ensembles testés.

En effet, lorsque la géométrie était du type $\mathcal{V}_2^x$, nous avions testé 6 ensembles, alors que dans le cas suivant, 9 ensembles de paramètres étaient testés. Il en résulte qu'un rang moyen de 2.5 dans le premier cas, pourrait signifier 3.5 ou 4 dans le second cas. En conséquence, la comparaison des deux premiers ensembles précités permet d'établir la hiérarchie suivante:

- Meilleur modèle : $\mathcal{V}$ du type $\mathcal{V}_2^x$, variance expérimentale égale à 145 nT^2.
- Modèle acceptable : $\mathcal{V}$ du type $\mathcal{V}_3^x$, variance expérimentale égale à 232 nT^2.

On attribuera la différence à l'effet "sautant" plus marqué de $\mathcal{V}_3^x$.

On notera que dans les deux cas de figure, c'est une FAI-2 qui a été retenue pour modéliser l'invariant spatial Y(x), et ceci pour des géométries de $\mathcal{V}$ très différentes. Cette permanence, à travers les géométries, du degré "k" de la FAI est une preuve expérimentale de la robustesse de ce paramètre pour ce jeu de données.

On rappellera enfin que mis à part le modèle (non significatif) lié à la configuration $\mathcal{V}_4^{x\,t}$, la période $T_2 = 12^h$ a dû être abandonnée car non détectable.

D.3 Inférence et validation de K(h)

Trois ensembles de paramètres s'étant dégagés des analyses précédentes, il reste maintenant à leur associer des modèles de covariances généralisées.

On rappelle que pour chacun de ces 3 ensembles, plusieurs modèles de covariance seront testés et ajustés, et le choix s'opèrera selon deux critères très différents :

- Le critère du "Jack-knife" qui représente approximativement le rapport d'une variance expérimentale sur une variance théorique. Il doit être aussi proche de 1 que possible. Les tests s'établissent sur les voisinages étendus puis décomposés. Les géométries résultantes peuvent alors différer sensiblement des géométries fixées.
- Le critère des variances expérimentales où l'opérateur d'estimation sera testé "in-situ" et sur des voisinages possédant exactement la géométrie demandée .

Il conviendra de comparer entre eux ces critères, leurs conflits ou leurs cohérences.

On rappelle également que le modèle suivant de CG sera utilisé pour modéliser K(h) :

$$K(h) = \underbrace{a_0 \text{ pépite}(|h|)}_{\text{effet de pépite}} + \underbrace{a_1 |h|}_{\text{linéaire}} + \underbrace{a_2 h^2 \text{ Log } |h|}_{\text{spline}} + \underbrace{a_3 |h|^3}_{\text{cubique}} \qquad (1)$$

avec : h , distance spatiale

Notre travail consistera à déterminer les structures à garder dans cette formule (1), à calculer les coefficients a_i, et enfin à évaluer les performances de ce modèle.

Inférence de K(h) conditionnellement à $\{\mathcal{V}_2^x \; ; k=2 \; ; 24^h\}$:

L'ensemble de paramètres $\{\mathcal{V}_2^x \; ; k=2 ; T=24^h\}$, étant fixé, il reste à fixer le modèle K(h) optimal pour ces paramètres.

Après avoir testé toutes les configurations possibles de (1), une pré-sélection des trois modèles suivants de covariance généralisée a été faite selon le critère du Jack-knife :

- $K_1^1(h) = a_1 |h|$
- $K_2^1(h) = a'_1 |h| + a'_2 |h|^2 \text{ Log}(|h|)$
- $K_3^1(h) = a''_0 \text{ pépite}(h) + a''_2 |h|^2 \text{ Log}(|h|)$

On rappelle que pour cette pré-sélection, le calcul des coefficients ainsi que les Jack-knife résultants sont établis sur des géométries sensiblement différentes de celle spécifiée (ici $\mathcal{V}_2^x$).

modèle retenu

		Modèle pour K(h)					
Rang	**Jack-knife**	**pépite**	**linéaire** $a_1\|h\|$	**spline** $a_2h^2\log\|h\|$	**cubique** $a_3\|h\|^3$	**$E(Z-Z_*)^2$** nT^2	**Rang**
1	*0.700*		●			*0.188*	*1*
2	*0.300*		●	●		*0.262*	*2*
3	*0.299*	●		●		*0.273*	*3*

INFERER (boucle-k et boucle-K) — **VALIDER (boucle-K*)**

Figure d.4 : Inférence de la covariance généralisée K (h) (conditionnellement au premier ensemble de paramètres)

Les trois modèles précités ont été classés selon leur Jack-knife qui est respectivement 0.7, 0.3 et 0.229 (figure d.4). L'examen de ce premier critère permet de constater que bien que sa valeur (pour le premier modèle) soit très éloignée de 1, elle représente cependant de loin le meilleur résultat (on passe de 0.7 à 0.3). Une telle différence permettrait déjà d'éconduire les autres modèles.

La "boucle-K*" de validation de ces trois modèles donne les variances expérimentales qui figurent en d.4. On y constate que globalement, les variances sont très faibles (puisque toutes inférieures à 0.3 nT^2).De plus, on constate que les 2 critères de Jack-knife et de variance expérimentale sont cohérents puisqu'ils ordonnent les modèles de la même façon. En conséquence, conditionnellement à l'ensemble $\{ V_2^x ; k=2 ; T_1 = 24^h \}$, c'est le modèle suivant de covariance généralisée qui est retenu :

$K^1(h) = a_1 |h|$ avec $a_1 = -1.02$

Inférence de K(h) conditionnellement à $\{ \gamma_3^x ; k=2 ; 24^h \}$:

L'ensemble de paramètres $\{ \gamma_3^x ; k=2 ; T=24^h \}$, étant fixé, il reste à fixer le modèle K(h) optimal pour ces paramètres.

Après avoir testé toutes les configurations possibles de la formule (1), une présélection de trois modèles de covariance généralisée a été faite selon le critère du Jack-knife :

- $K_1^2(h) = a_\emptyset$ pépite(h) $+ a_3 |h|^3$
- $K_2^2(h) = a'_3 |h|^3$
- $K_3^2(h) = a''_2 |h|^2 \text{ Log}(|h|)$

Ces trois modèles ont été classés selon des Jack-knife respectivement égaux à 1.026 , 0.960 et 1.239 (figure d.5).

modèle retenu

		Modèle pour K(h)					
Rang	**Jack-knife**	**pépite**	**linéaire** $a_1\|h\|$	**spline** $a_2h^2 \log \|h\|$	**cubique** $a_3\|h\|^3$	**$E(Z-Z_*)^2$** nT^2	**Rang**
1	*1.026*	●			●	*0.810*	*2*
2	*0.960*				●	*2.045*	*3*
3	*1.239*			●		*0.295*	*1*

INFERER (boucle-k et boucle-K)

VALIDER (boucle-K*)

Figure d.5 : Inférence de la covariance généralisée K (h) (conditionnellement au deuxième ensemble de paramètres)

L'examen de ce premier critère permet de constater que les deux premières valeurs sont très proches de 1, et peu différentes l'une de l'autre, ce qui ne permet pas de choisir entre ces deux premiers modèles.

La "boucle-K*" de validation directe de ces trois modèles a permis d'établir les variances expérimentales qui figurent en d.5. On y constate que globalement, ces variances sont supérieures à celles obtenues pour la géométrie du type $\mathcal{V}_2^x$, mais toujours extrêmement faibles, compte tenu des variances expérimentales rencontrées dans la boucle-k-$\{f_i\}$.

Comme ces variances ne sont pas sélectives, c'est le Jack-knife qui prévaudra. En conséquence, conditionnellement à l'ensemble $\{\mathcal{V}_3^x\ ,\ k=2,\ T=24^h\}$ de paramètres, nous adopterons le modèle suivant :

$K^2(h) = a_\emptyset$ pépite$(h) + a_3\ |h|^3$ avec $a_\emptyset = 2.0190$ et $a_3 = 0.1672$

Inférence de K(h) cond. à $\{\mathcal{V}_4^{x\,t}\ ;\ k=2\ ;\ 24^h, 12^h\}$:

On rappelle que cette géométrie spatio-temporelle fut rejetée en raison du conflit total qui en résultait dans la boucle précédente.. C'est donc uniquement par curiosité qu'elle continue de figurer dans ce cas d'étude.

L'ensemble de paramètres $\{\mathcal{V}_4^{x\,t}\ ;\ k=2\ ;\ T=24^h, 12^h\}_{prior}$ fixé, il faut déterminer un modèle K(h) optimal. Après avoir testé toutes les configurations possibles de (1) une pré-sélection de trois modèles de covariance généralisée a été faite selon le critère du Jack-knife:

- $K_1^3(h) = a_2\ |h|^2\ \mathrm{Log}(|h|)$
- $K_2^3(h) = a'_1\ |h| + a'_3|h|^3$
- $K_3^3(h) = a''_\emptyset$ pépite$(h) + a''_3\ |h|^3$

Ces trois modèles ont été classés selon leurs Jack-knife respectivement égaux à 0.862, 0.607 et 0.527. On note que globalement, ces valeurs sont intermédiaires entre celles des géométries $\mathcal{V}_2^x$ et $\mathcal{V}_3^x$. De plus, ce premier critère donne un avantage sensible au premier modèle (figure d.6).

modèle retenu

		Modèle pour K(h)					
Rang	**Jack-knife**	**pépite**	**linéaire** $a_1\|h\|$	**spline** $a_2h^2\log\|h\|$	**cubique** $a_3\|h\|^3$	$E(Z-Z_*)^2$ nT^2	**Rang**
1	*0.862*			●		*0.616*	*1*
2	*0.607*		●		●	*0.673*	*2*
3	*0.527*	●		●	●	*1.114*	*3*

INFERER
(boucle-k et boucle-K)

VALIDER
(boucle-K*)

Figure d.6 : Inférence de la covariance généralisée K (h)
(conditionnellement au troisième ensemble de paramètres)

La "boucle-K*" de validation directe de ces trois modèles a conduit aux variances expérimentales de la figure d.6. On y constate que ces variances expérimentales se comportent d'une manière similaire aux Jack-knife puisque :

- globalement, elles offrent des valeurs intermédiaires entre les deux précédentes géométries,

• elles ordonnent les trois modèles de covariance selon la même hiérarchie.

La cohérence est totale entre ces deux critères, ce qui fait que conditionnellement à l'ensemble $\{ \mathcal{V}_4^{x\,t} ; k=2 ; T=24^h,12^h \}$ c'est le modèle suivant de covariance généralisée qui a été choisi :

$K^3(h) = a_2 |h|^2 \text{ Log}(|h|)$ avec $a_2 = 1.25$

On rappelle qu'outre le fait qu'il engendrait dans la "boucle-k" un conflit total, ce dernier modèle est basé sur un effectif statistique extrêmement réduit (35 voisinages seulement), et ceci à cause des contraintes temporelles rigoureuses et d'un diamètre spatial du voisinage qui est important.

Si on n'avait pas pris soin d'examiner dans le détail, les résultats de la "boucle-k", la cohérence obtenue à l'issue des boucles "K" et "K" pourrait faire croire (à tort) que ce modèle est bon.*
On voit donc l'intérêt d'être attentif aux résultats et tests, sous peine d'être orienté par la machine, vers un modèle faux, et faux partout (puisque cohérent).

D.4 Conclusions

Les modèles inférés :

Les analyses ont conduit aux trois modèles suivants :

- $\{ \mathcal{V}_2^x ; k = 2 ; T = 24^h ; K^1(h) \}$

  avec $K^1(h) = a_1 |h|$

- $\{ \mathcal{V}_3^x ; k = 2 ; T = 24^h ; K^2(h) \}$

  avec $K^2(h) = a_\beta$ pépite$(h) + a_3 |h|^3$

- $\{ \mathcal{V}_4^{x\,t} ; T = 24^h, 12^h ; k = 2 ; K^3(h) \}_{\text{prior}}$

  avec $K^3(h) = a_2 |h|^2 \text{ Log}(|h|)$

Dans le troisième modèle, on rappelle que l'indice "prior" existe pour rappeler que ce modèle n'a pas été inféré.

La "boucle-K*" n'est pas sélective :

Toutes les variances obtenues dans les "boucles-K*" sont extrêmement faibles, puisque de l'ordre de 1 nT^2, alors qu'elles étaient de l'ordre de 100 nT^2 (voire plus) dans la boucle-k-{fi}. Pourquoi? La raison en est simple : bien que les géométries de $\mathcal{V}$ soient contraintes à s'étendre spatialement (afin d'éviter les amas de points), des amas subsistent cependant, et ceci autour du centre du voisinage. Ainsi, une configuration du type $\mathcal{V}_2^x$ (où seuls quelques points sont contraints géométriquement), contient pour les points complémentaires, une grappe de points alignés sur le profil porteur du point ciblé par l'analyse. Or, les statistiques de ce deuxième critère (variance) sont établies à partir d'accroissements de la forme $(z(x, t) - z^*(x,t))$ où $z(x, t)$ a été supprimé de $\mathcal{V}$, puis estimé, ce qui engendre $z^*(x, t)$. Il en résulte que même si dans cette opération (effectuée sur une centaine de voisinages), $z(x, t)$ est supprimée de $\mathcal{V}$, il lui subsiste des points voisins (voire quasiment confondus) qui possèderont une influence considérable dans l'estimation. Le résultat est que $z(x, t)$ n'est que virtuellement évincé de $\mathcal{V}$ et subsiste dans les faits. Dès lors, la CLA-k-$\{f_i\}$ $\sum \lambda^\alpha Z(x_\alpha, t_\alpha)$ est équivalente à la seule valeur $z(x, t)$: aucun test n'est fait, aucune estimation n'est vraiment conduite, et la variance expérimentale (de l'erreur d'estimation) est quasiment nulle.

On peut s'étonner alors qu'une géométrie du type $\mathcal{V}_3^x$ où tous les points sont contraints, possède cette même caractéristique d'engendrer des variances expérimentales

anormalement faibles, puisque en théorie, le voisin le plus proche sera en moyenne à une distance égale à une demi-longueur de "pseudo" profil (9 km dans ce cas d'étude).
Supposons que Npoint éléments soient demandés pour constituer le voisinage. Comme parmi ceux-ci figure la cible de l'analyse (qui sera ultérieurement retirée pour l'estimation), c'est en fait un total de (Npoint + 1) éléments qui doit être construit. Ce point supplémentaire que nous n'avons pas su contrôler, fut choisi seulement selon un critère de proximité immédiate de la cible. Comme les données sont sur profil, c'est un point quasiment confondu avec la cible qui est pris. Il en résulte que ce qui est supprimé (la cible) est d'une certaine manière réinjecté dans l'ensemble des points sélectionnés, avec pour conséquence immédiate que les variances expérimentales ne représentent pas grand-chose. Nous expérimenterons une solution à ce problème dans le cas d'étude Mohns.

En conclusion, on privilégiera toujours le critère de Jack-knife si les variances expérimentales de la "boucle-K*" sont anormalement faibles.

E Détails de la modélisation de la campagne Seaperc

Dans ce qui suit, nous donnons le détail de l'inférence des modèles qui ont permis le filtrage des perturbations dans le chapitre 4. Une grande attention est portée aux géométries des voisinages, et ceci au détriment du détail de l'inférence des paramètres des modèles.

Pour mener à bien la modélisation de ce PTP, nous adopterons comme pour Krakatao 2 types de séquencement de la route du bateau :

- un premier pour lequel la longueur (Δcode) sera de 400 minutes et qui conduira à une géométrie pour $\mathcal{V}$ du type $\mathcal{V}_2^x$ (figure e.1)
- un deuxième dont la longueur Δcode est de 1 heure, et qui conduira à des géométries de type $\mathcal{V}_3^x$ et $\mathcal{V}_4^{x\,t}$ (figure e.2)

E.1 Modèle 1 : $\mathcal{V}$ du type $\mathcal{V}_2^x$, T=24h

La décomposition de la route est celle des profils "lâches" (Δ code = 400 mn). La population de $\mathcal{V}$ a été fixée à Npoint=20 éléments dont 7 devant appartenir à 7 profils différents (Ncode=7). La configuration résultante est donc celle du type $\mathcal{V}_2^x$ (spatial-anisotrope).

La figure e.3 représente dans l'espace les configurations résultantes de la géométrie $\mathcal{V}_2^x$ lorsqu'elle est appliquée à ces données. On y présente 3 cas de figure, selon l'implantation du centre du voisinage : la bordure Nord, le centre, et la bordure Sud de la zone d'étude. Leur analyse montre :

- Une forte anisotropie Nord-Est, Sud-Ouest qui fait que la géométrie des points est sensiblement une droite. Cette configuration peut rendre le système d'analyse inapplicable lorsque k > $\emptyset$
- Un fort regroupement des points non conditionnés par les contraintes d'extension de $\mathcal{V}$ sur les "pseudo-profils".
  En effet, parmi les 20 points qui constituent $\mathcal{V}$, 7 seulement sont conditionnés par les spécifications Δ code et Ncode, et ce sont ceux-là qui donnent à $\mathcal{V}$ sa géométrie anisotrope. Les 13 points complémentaires sont pris au plus près du point ciblé par l'analyse et constituent de ce fait un amas de points puisque 300 m seulement séparent 2 points successifs.

La figure e.4 représente le pendant temporel de la figure e.1 puisqu'elle contient l'échantillonnage temporel résultant des configurations spatiales de type $\mathcal{V}_2^x$, et ceci en regard des oscillations de la série Huancayo qui représente pour nous une image acceptable de la réalité des perturbations dans z(x,t).

On note qu'il est possible d'associer à $\mathcal{V}_2^x$ un pas temporel d'échantillonnage proche de 12^h . Ici encore, ce sont les points conditionnés par les codes qui donnent à $\mathcal{V}$ ses caractéristiques, les autres constituant des amas. Comme à cette échelle, les distances entre ces points non conditionnés peuvent être considérées comme nulles, tout se passe en terme d'échantillonnage comme si ces derniers échantillons n'existaient pas. Cet échantillonnage ne respecte donc pas les contraintes du théorème d'échantillonnage (ref [TdS-5]) puisque le pas est égal à $\frac{T}{2}$ .De plus, il n'entre pas dans les deux configurations possibles d'échantillonnage d'un phénomène périodique puisque ce pas est un diviseur de T. En conséquence, lorsque la fenêtre d'analyse se déplace dans l'espace (et de ce fait dans le temps), cette configuration échantillonne tantôt les extréma de D(t) (parties "A" et "C" de la figure), tantôt une zone où D(t) ne présente pas d'activité (partie "B" de la figure). Dans ce dernier cas, l'ajustement par moindre carré (boucle-k-$\{f_i\}$) ou la validation des modèles (boucle-K*) donnera des valeurs arbitraires puisque des points ainsi alignés sont incohérents avec les fonctions trigonométriques utilisées pour l'ajustement de la composante temporelle de z(x,t). Dans le meilleur des cas, le système d'analyse ne détectera pas de tendance temporelle, et dans le pire des cas, l'estimateur utilisé dans la boucle de validation sera instable.

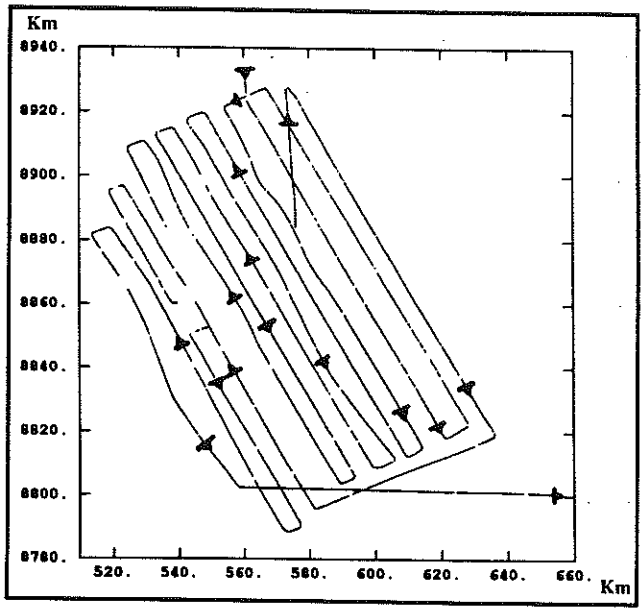


Figure e.1 : Séquencement de la route du bateau en "pseudo–profils" lâches (campagne Seaperc)

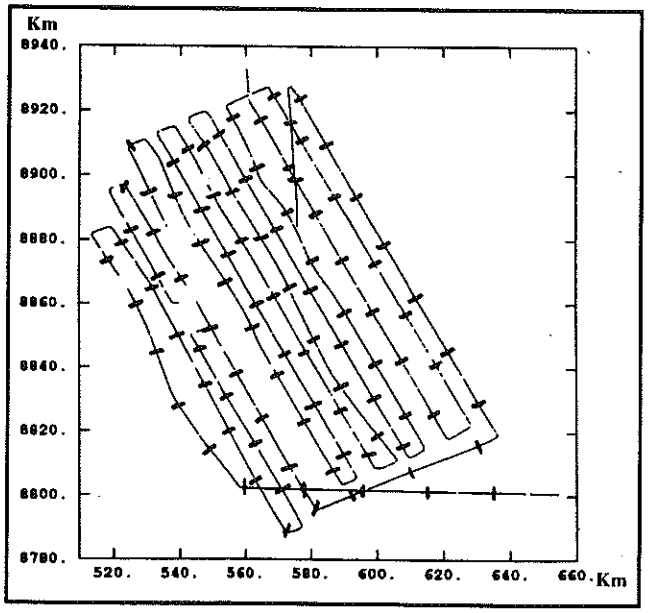


Figure e.2 : Séquencement de la route du bateau en "pseudo–profils" courts

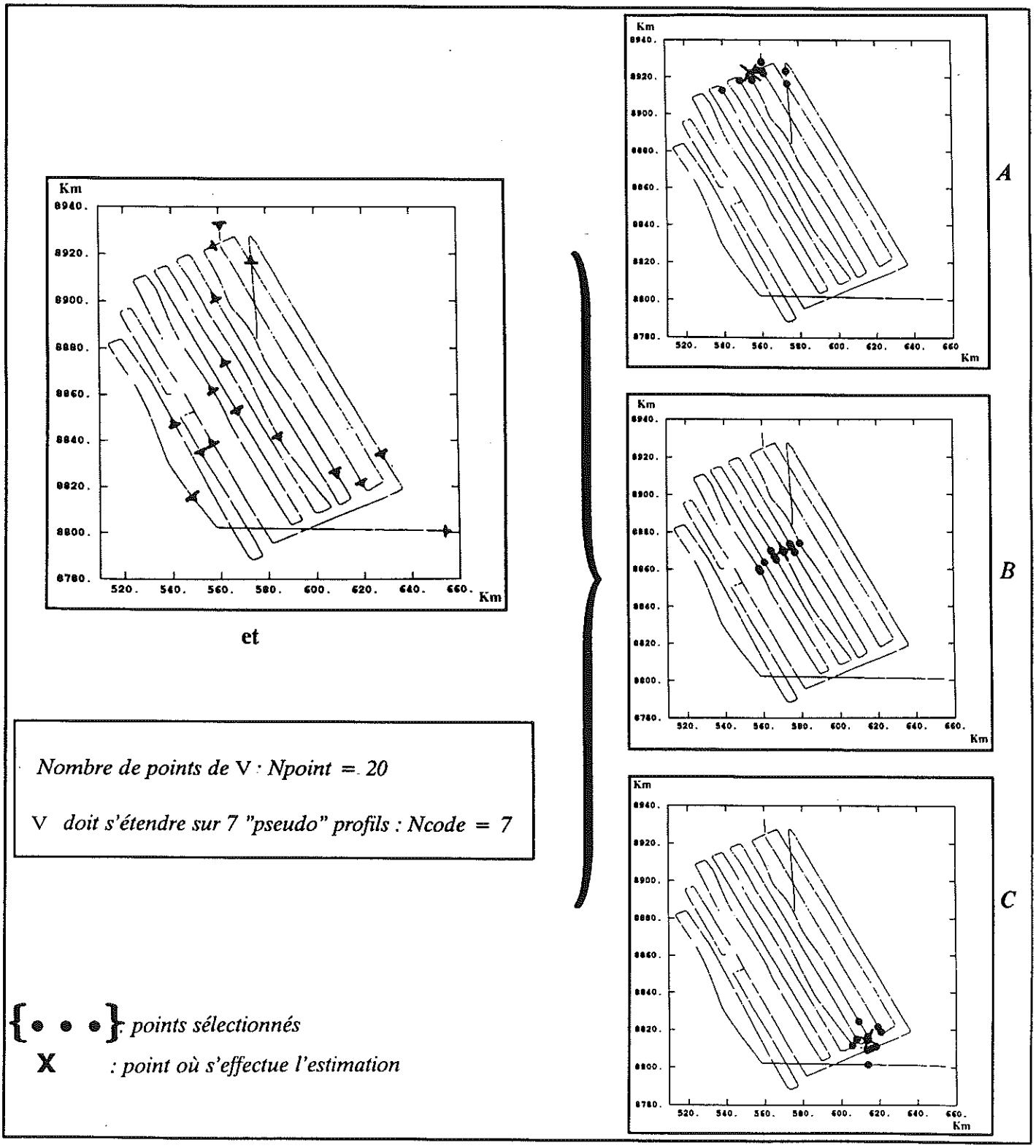


Figure e.3 : Configuration spatiale des échantillons sélectionnés lorsque la géométrie est du type γ_2^x : "spatiale anisotrope"

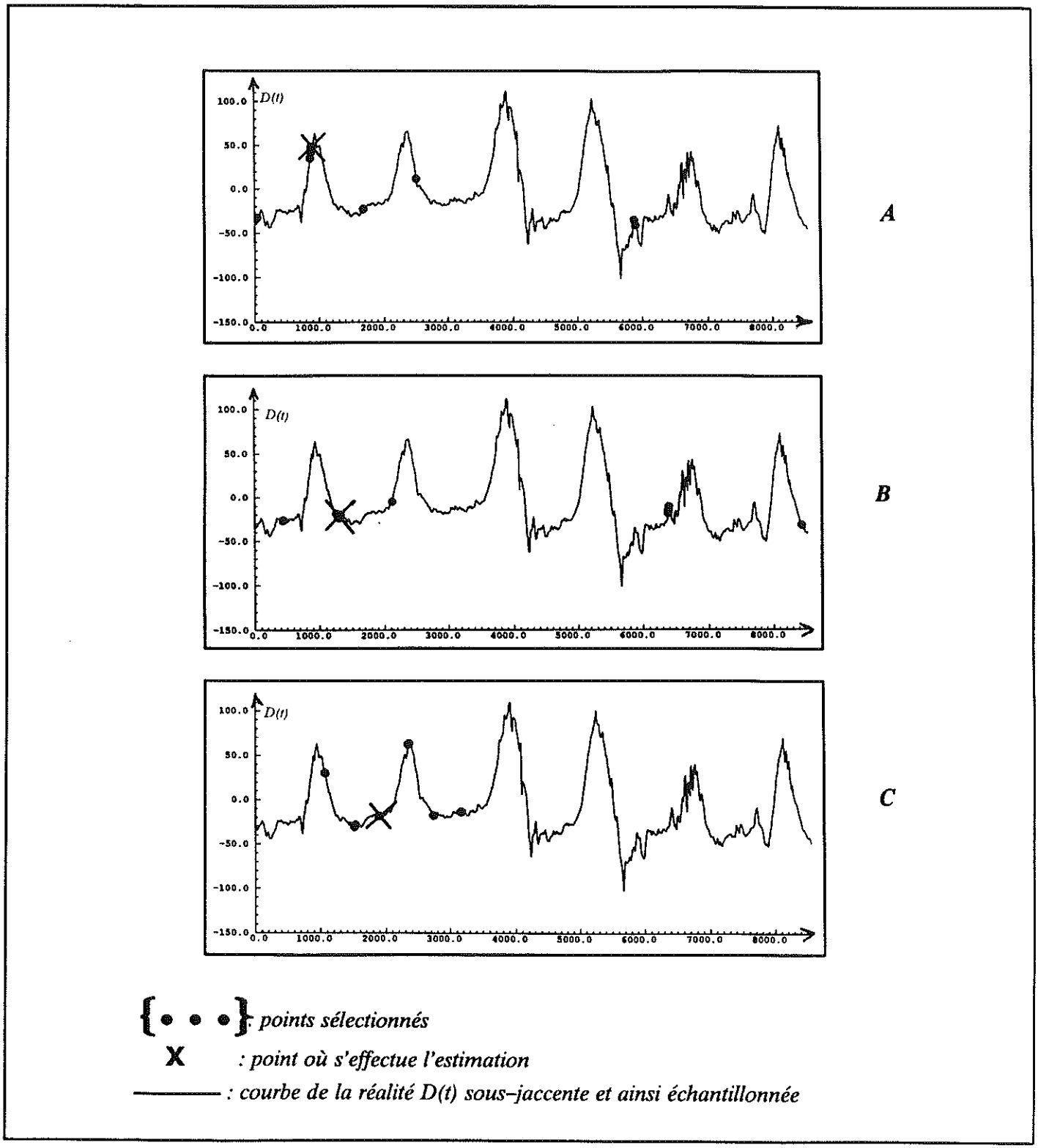


Figure e.4 : Configuration temporelle des échantillons sélectionnés (géométrie du type γ_2^x : "spatiale anisotrope")

La boucle-k-$\{f_i\}$ nous a mis en présence d'un conflit total entre le critère stable (rang moyen) et le critère de variance expérimentale. Nous avons donc dû imposer la période $T = 24^h$ qui de ce fait devient un paramètre "a priori".

L'inférence de "k" conditionnellement à $T_{prior} = 24^h$ et à la géométrie $\mathcal{V}_2^x$ a donné l'optimalité pour $k = 1$.

L'ensemble de paramètres $\{ \mathcal{V}_2^x ; k = 1 ; T_{prior} = 24^h \}$, étant fixé, le modèle K(h) optimal pour ces paramètres est :

$K^1(h) = a_2 \, h^2 \, Log(|h|)$ avec $a_2 = 10.62$

E.2 Modèle 2 : $\mathcal{V}$ du type $\mathcal{V}_2^x$, $T = 24^h$ et 12^h

Nous allons maintenant inférer un modèle contenant également la période $T = 12^h$, bien qu'en vertu de l'expérience précédente, on sache que 24^h et 12^h sont ici inaccessibles. Les mêmes paramètres géométriques que ceux de l'expérience 1 sont utilisés, à savoir Npoint = 20, Δ code = 400 minutes, et Ncode = 7, de sorte que la géométrie résultante sera ici encore du type $\mathcal{V}_2^x$.

Comme précédemment, les périodes 24^h et 12^h ont été imposées.
L'inférence de "k" conditionnellement à $T_{prior} = 24^h, 12^h$ et à la géométrie $\mathcal{V}_2^x$ a donné l'optimalité pour $k = 1$.
En ce qui concerne le modèle K(h) et après avoir testé toutes les configurations possibles de la formule (1), c'est finalement le modèle suivant de CG qui à été retenu :

$K^2(h) = a_2 \, h^2 \, Log(|h|)$ avec $a_2 = 6.188$

E.3 Modèle 3 : $\mathcal{V}$ du type $\mathcal{V}_3^x$, $T = 24^h$ et 12^h

Cette troisième expérience se différencie de la précédente par la géométrie utilisée qui devient du type $\mathcal{V}_3^x$. La décomposition de la route est celle des pseudo-profils tels que Δ code = 1^h. La population de $\mathcal{V}$ a été fixée à Npoint = 30 éléments qui doivent tous appartenir à 30 pseudo- profils différents (Ncode = 30). La configuration résultante est donc celle du type $\mathcal{V}_3^x$ (spatial-anisotrope).

La figure e.5 représente les géométries résultantes de la configuration $\mathcal{V}_3^x$ lorsqu'elle est appliquée à ces données. Cette figure illustre bien les discussions du chapitre 3 relatives à cette géométrie, à savoir que si elle est isotrope, elle sera "sautante" puisque ce sont en majorité les points frontières qui sont pris; de sorte que $\mathcal{V}$ ne se renouvellera pas continûment mais par sauts, lorsque la fenêtre d'analyse se déplacera sur le champ d'analyse.
On constate de plus que le diamètre spatial de la fenêtre s'est considérablement étendu : il est passé de 40 km (géométrie précédente) à 60 km dans le cas présent.

La géométrie temporelle résultante de $\mathcal{V}_3^x$ est représentée sur les figures e.6. On y constate que si les points échantillonnent mieux la tendance temporelle (partie "A" et "C"), certains cas de figure peuvent se présenter où comme précédemment , les points sont sensiblement alignés et ne permettent donc pas une inférence correcte des périodes. Bien qu'occasionnelles, ces configurations critiques interviendront périodiquement lorsque la fenêtre se déplacera le long des profils. Elles peuvent donc nuire à l'analyse et à la validation des modèles.

Cette géométrie présente un meilleur échantillonnage temporel et de ce fait, les périodes $T = 24^h$ et 12^h ont pu être inférées, tant par le critère de rang optimal que par les variances expérimentales. L'optimalité (conditionnelle) est ici encore obtenue avec $k = 1$.

L'ensemble de paramètres $\{ \mathcal{V}_3^x ; k = 1 ; T = 24^h, 12^h \}$ étant fixé, le modèle K(h) optimal pour ces paramètres est :

$K^3(h) = a_\emptyset$ pépite(h) $+ a_2 \, h^2 \, Log(|h|)$ avec $a_\emptyset = 0.213$ et $a_2 = 6.16$

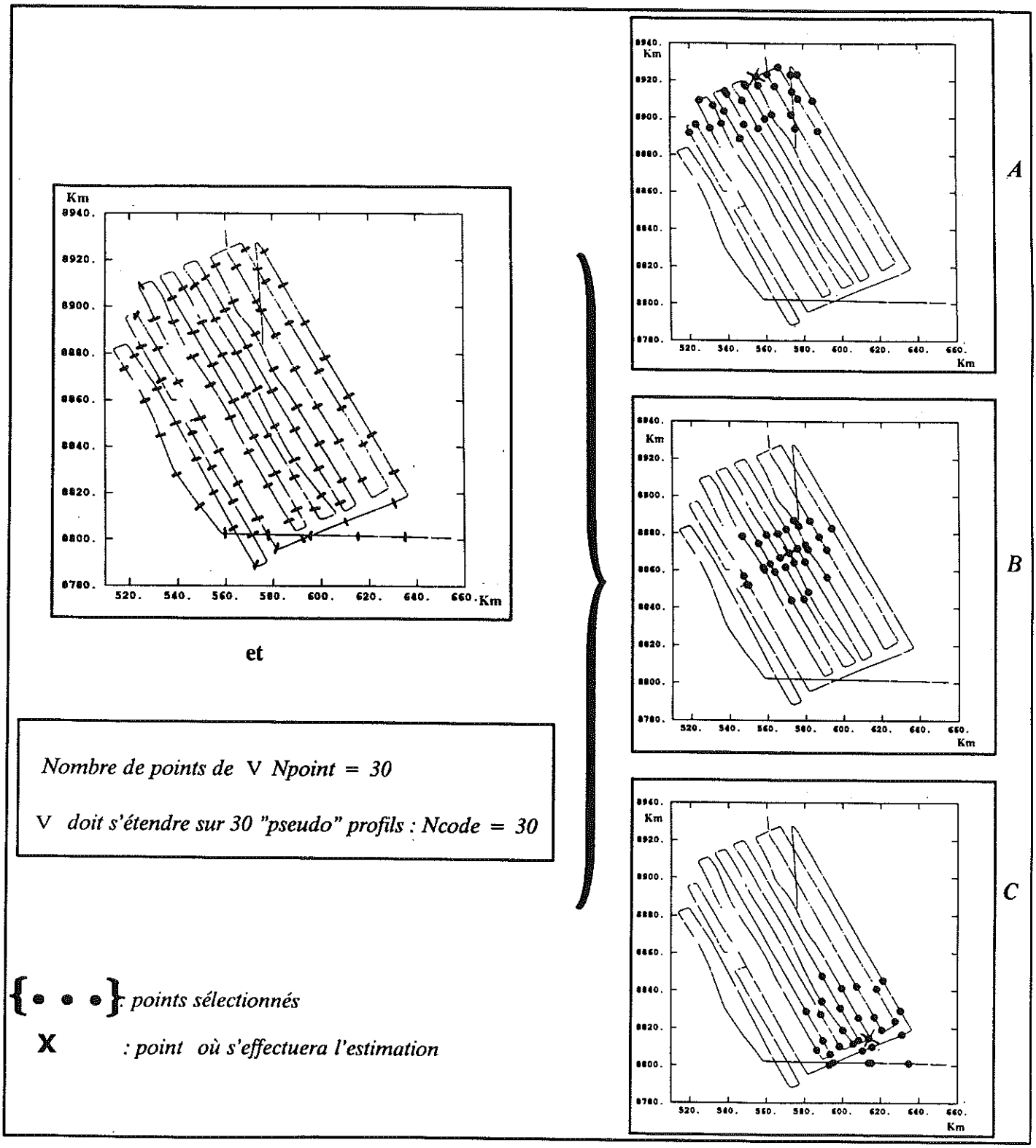
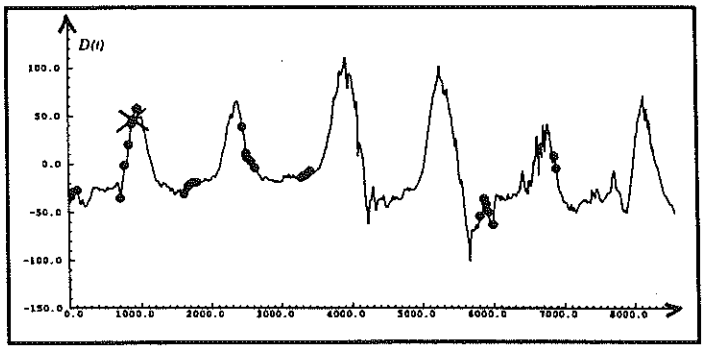
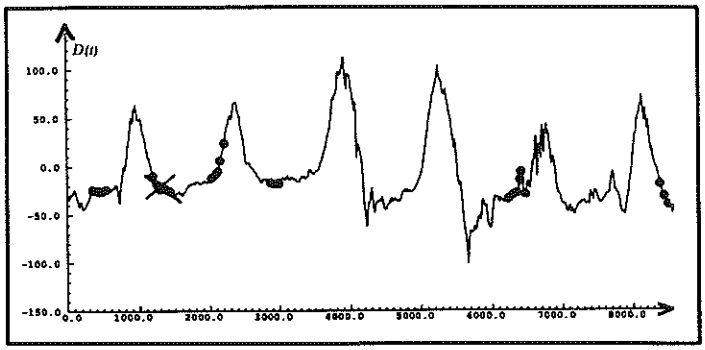


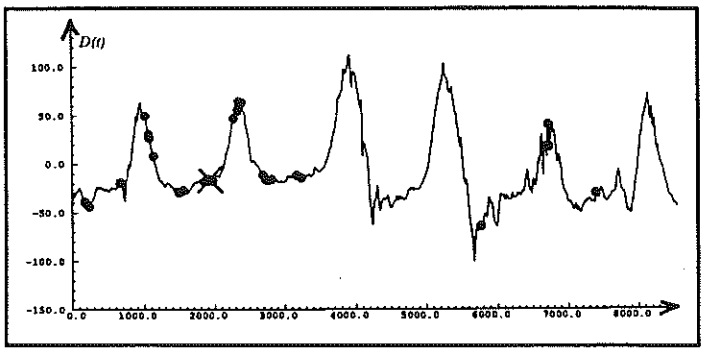
Figure e.5 : Configuration spatiale des échantillons (géométrie du type φ_3^x : "spatiale–isotrope–sautante")



A



B



C

{• • •} *points sélectionnés*

X *: point où s'effectuera l'estimation*

——— *: courbe de la réalité D(t) sous-jaccente et ainsi échantillonnée*

Figure e.6 : Configuration temporelle des échantillons (géométrie du type $\mathcal{V}_3^X$: "spatiale-isotrope-sautante")

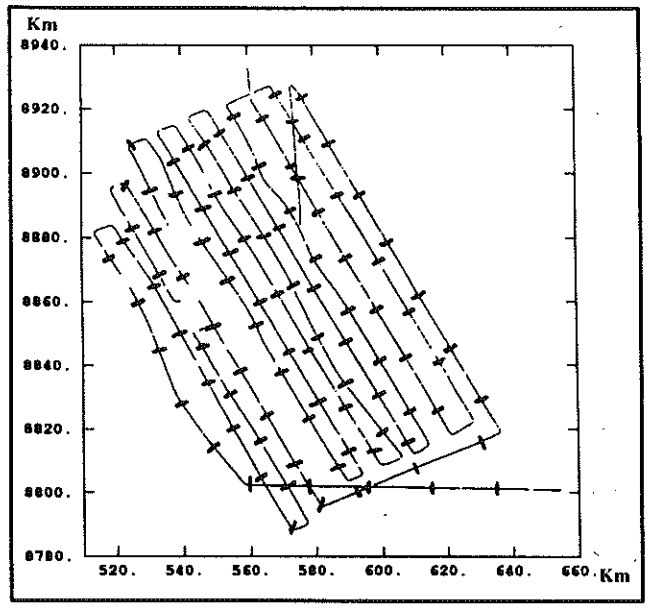
E.4 Modèle 4 : $\mathcal{V}$ du type $\mathcal{V}_4^{x\,t}$, T=24^h et 12 ^{h}

Notre objectif est de tenter de garder les caractéristiques spatiales de $\mathcal{V}_3^x$ (isotropie), tout en atténuant son effet "sautant" par des contraintes temporelles "glissantes" avec le temps ciblé par l'analyse.
Le séquencement de la route est le même que précédemment (Δcode = 60 minutes), le nombre Npoint d'échantillons contenus dans $\mathcal{V}$ est porté à 30 points, dont seuls 20 points sont contraints à s'étendre sur 20 pseudo-profils différents (variable Ncode). Les 10 points complémentaires seront ici contraints à encadrer (temporellement) t_x (temps ciblé par l'analyse) à maille régulière, par pas de 60 minutes, et symétriquement autour de t_x. On rappelle que cette partie du voisinage est glissante dans le temps. L'effet escompté est que cette partie temporelle glissante atténue l'effet sautant de la partie spatiale $\mathcal{V}_3^x$ et échantillonne mieux le phénomène temporel.

La figure e.7 montre la configuration de l'échantillonnage lorsque cette géométrie est appliquée aux données. On y retrouve une configuration spatiale désormais connue du type $\mathcal{V}_3^x$ moins étendue que précédemment (Ncode = 20, seulement) à laquelle s'ajoutent 10 points à contraintes temporelles qui étendent considérablement $\mathcal{V}$ et lui font perdre l'isotropie de la géométrie $\mathcal{V}_3^x$

La géométrie temporelle résultante apparaît sur la figure e.8 qui montre que les 10 points contraints temporellement permettent un meilleur échantillonnage de la fonction, même dans les configurations critiques (partie "B" de la figure).

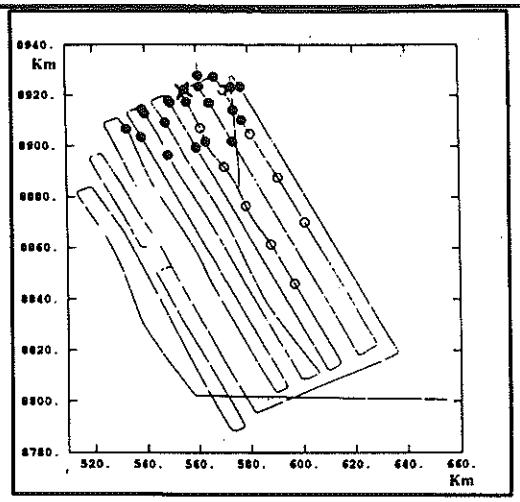
A cause des problèmes d'inférence rencontrés précédemment et liés à cette géométrie, nous avons choisi de garder les mêmes caractéristiques que le modèle 4 pour k et K(h), de sorte que ce modèle est intégralement "a priori".



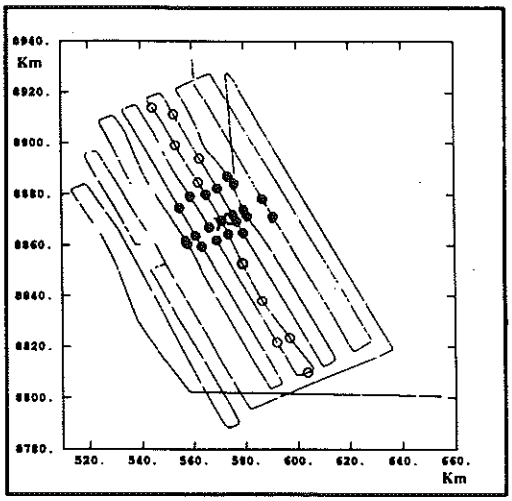
et

Nombre de points de V *Npoint = 30*

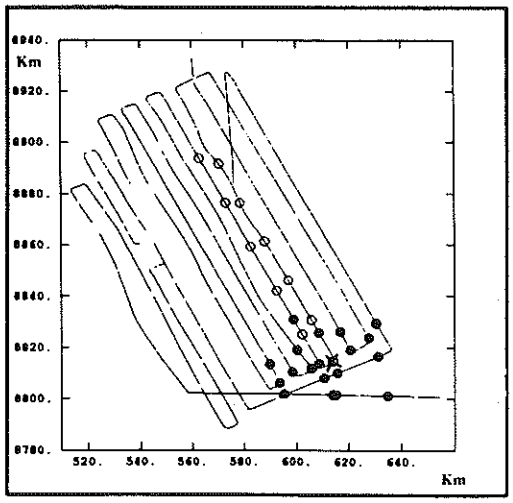
V *doit s'étendre spatialement sur 20 "pseudo" profils et sur 10 heures autour de* t_x



A



B



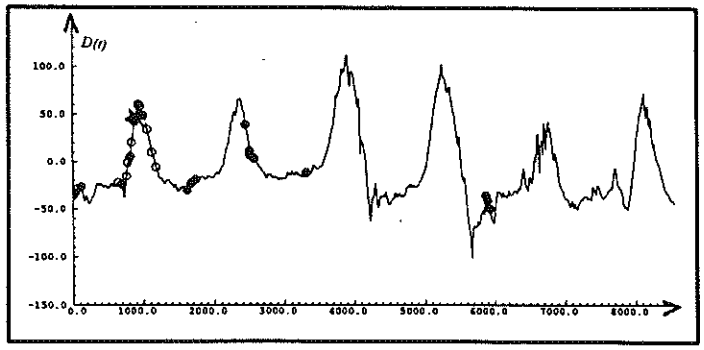
C

X : *point où s'effectuera l'estimation*

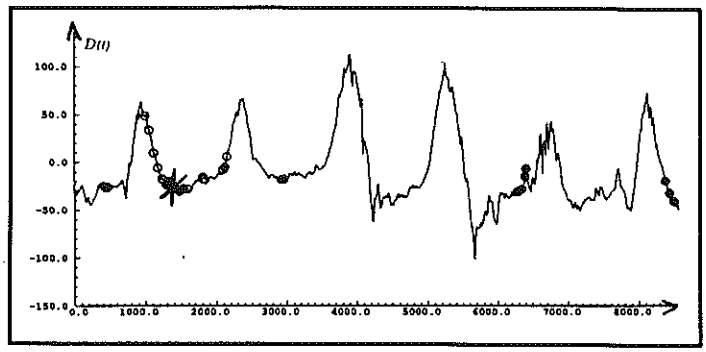
{ ○ ○ ○ } : *points sélectionnés sous contraintes temporelles*

{ ● ● ● } : *points sélectionnés sous contraintes spatiales*

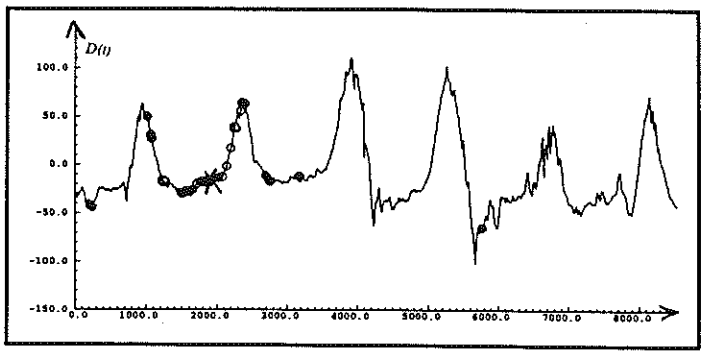
Figure e.7 : Configuration spatiale des échantillons (géométrie du type $\varphi_4^{x\,t}$: "spatio-temporelle")



A



B



C

{ ○ ○ ○ } : *points sélectionnés sous contraintes temporelles*

{ ● ● ● } *points sélectionnés sous contraintes spatiales*

X : *point où s'effectuera l'estimation*

——— : *courbe de la réalité D(t) sous-jaccente et ainsi échantillonnée*

Figure e.8 : Configuration temporelle des échantillons (géométrie du type $\varphi_4^{X\,t}$: "spatio-temporelle")

E.5 Modèle 5 : $\mathcal{V}$ du type $\mathcal{V}_4^{x\,t}$, FAI-$\emptyset$ forcée, T=24h, 12h

Dans cette dernière modélisation, nous pousserons l'arbitraire jusqu'à imposer le degré de la FAI en posant $k_{prior} = \emptyset$.

Le voisinage adopté ici, est celui du type $\mathcal{V}_4^{x\,t}$ précédemment utilisé, pour le même séquencement en pseudo-profils de longueur $\Delta code = 60^{mn}$, avec une population de 30 points (Npoint) dont 20 sur 20 pseudo-profils différents (Ncode) et 10 à maille régulière centrée sur la cible t_x, par pas de 1 heure. Cette géométrie n'a pas permis d'inférer la présence de périodes temporelles. La situation est donc comparable à celle de Krakatao. En conséquence et pour cette géométrie, nous nous sommes imposé $T_{prior} = 24^h , 12^h$.

L'ensemble de paramètres $\{ \mathcal{V}_4^{x\,t} , k = \emptyset , T = 24^h , 12^h \}_{prior}$ étant fixé, le modèle K(h) optimal pour ces paramètres est :

$K^5(h) = a_1 |h|$ avec $a_1 = -3.672$

E.6 Conclusions

Les modèles inférés :

Les analyses ont permis d'aboutir à 5 ensembles de paramètres dont certains arbitraires seront indexés par "prior" :

- $\{ \mathcal{V}_2^x ; k = 1 ; T_{prior} = 24^h ; K^1(h) \}$

  avec $K^1(h) = a_2 \, h^2 \, Log(|h|)$

- $\{ \mathcal{V}_2^x ; k = 1 ; T_{prior} = 24^h, 12^h ; K^2(h) \}$

  avec $K^2(h) = a_2 \, h^2 \, Log(|h|)$

- $\{ \mathcal{V}_3^x ; k = 1 ; T = 24^h, 12^h ; K^3(h) \}$

  avec $K^3(h) = a_\emptyset$ pépite(h) $+ a_2 \, h^2 \, Log(|h|)$

- $\{ \mathcal{V}_4^{x\,t} ; k_{prior} = 1 ; T_{prior} = 24^h, 12^h ; K^{prior}(h) = K^3 \}_{prior}$

  avec $K^4(h) = K^{prior}(h) = K^3(h)$

- $\{ \mathcal{V}_4^{x\,t} ; k_{prior} = \emptyset ; T_{prior} = 24^h, 12^h ; K^5(h) \}_{prior}$

  avec $K^5(h) = a_1 |h|$

Robustesse du degré "k" de la FAI :

Parmi les 5 ensembles disponibles, seuls les trois premiers sont tels que $k \neq k_{prior}$. Dans chacun de ces 3 premiers ensembles, on constate que k=1, et ceci pour des géométries très différentes de voisinage. Ce fait confirme la robustesse de ce paramètre pour le jeu de données Seaperc. On rappelle que cette conclusion était vraie également pour le jeu de données Krakatao.

Robustesse de K(h) :

Examinons à nouveau les trois premiers ensembles .

On constate que les modèles de CG inférés sont tous du type spline et ceci une fois de plus malgré des géométries très différentes dans les deux derniers ensembles.

Comparativement à Krakatao, un fait nouveau apparaît donc ici : la robustesse de K(h) pour ce jeu de données. On notera dans le troisième l'apparition d'un effet de pépite. On attribuera celui-ci à l'effet "sautant" plus marqué pour $\mathcal{V}_3^x$ que pour $\mathcal{V}_2^x$. On notera de plus que le passage du modèle 1 au 2 (ie l'introduction de $T = 12^h$ dans le système d'analyse) a entraîné la réduction à $a_2 = 6.188$ d'un coefficient pour K(h) qui dans le premier cas valait 10.62 . Pour le troisième modèle, ce coefficient garde une valeur comparable à celle associée au deuxième modèle, puisqu'il vaut 6.16 . Or, le troisième ensemble contient également les périodes 24^h et 12^h .

Ces faits semblent donc indiquer 2 choses essentielles :

- Dans le modèle 1, K(h) prend en compte une partie de l'énergie temporelle liée à $T = 12^h$, et dont on sait qu'elle existe dans z(x,t) (cf analyse de la série Huancayo, chap. 2).
- Lorsque les ensembles contiennent les périodes 24^h et 12^h , K(h) est extrêmement robuste vis à vis de la géométrie.

Difficultés de détection de $\{f_i\}$:

Si nous réexaminons ces 5 ensembles de paramètres, nous constatons que seul le troisième modèle contient des périodes non imposées. Ce fait illustre 2 points essentiels :

- Seule la configuration $\mathcal{V}_3^x$ permet d'appréhender convenablement les périodes temporelles.
- Il y a incompatibilité entre un bon échantillonnage spatial et un bon échantillonnage temporel, et ceci essentiellement parce que le temps et l'espace sont pour ce type de PTP, liés par la vitesse du bateau.

Cet état de fait condamne une configuration de type $\mathcal{V}_4^{x\,t}$ parce qu'elle détruit la configuration spatialement acceptable $\mathcal{V}_3^x$. Si ces circonstances rendent difficile la mise en oeuvre de la modélisation, elles n'en condamnent pas pour autant la méthode. Nous sommes simplement dans une situation délicate, et les inférences doivent être menées minutieusement.

L'a priori de certains des paramètres, ne doit pas heurter. Il est justifié ici par la connaissance physique que nous avons des phénomènes sous-jacents.

Cet a priori se verra ultérieurement confirmé ou infirmé lors de la phase finale de la chaîne méthodologique : l'estimation . C'est seulement à ce niveau que les conclusions pourront être définitives.

F Détails de la modélisation de la campagne Mohns

Dans ce qui suit, nous donnons le détail de l'inférence des modèles qui ont permis le filtrage des perturbations dans le chapitre 4. Une grande attention est portée à l'inférence des paramètres. On rappelle que par sa localisation boréale, les fluctuations temporelles qui affectent ces mesures, ont peu de chance d'être à "tendance" périodique, ce que confirment les trois jours d'enregistrement sur zone de perturbations.

F.1 Séquencement en "pseudo-profils"

Pour mener à bien l'analyse et la modélisation de ces données, nous adopterons comme pour Krakatao et Seaperc 2 types de séquencement de la route du bateau :

- Un premier pour lequel la longueur (Δcode) sera de 400 minutes et qui conduira à une géométrie pour $\mathcal{V}$ du type $\mathcal{V}_2^x$.
- Un deuxième dont la longueur Δcode est de 1 heure, et qui conduira à des géométries de type $\mathcal{V}_3^x$ et $\mathcal{V}_4^{x\,t}$.

Mais à la différence de ce qui a été fait jusqu'à présent, nous effectuerons une présélection arbitraire d'une donnée sur 10, afin d'augmenter la distance qui sépare deux échantillons successifs d'un même profil. Il en résultera que ces échantillons seront distants de 10 minutes de navigation, soit approximativement 3 km. L'objectif est de rendre les boucles K et K* plus sélectives . On rappelle en effet que les inférences dans Seaperc et Krakatao étaient biaisées par les amas de points qui donnaient des variances expérimentales extrêmement faibles, puisque significatives seulement de la variabilité moyenne entre 2 points distants de 300 mètres. En adoptant cette présélection, on supprime ces amas de points, et on espère de ce fait obtenir des résultats significatifs de tous les points de $\mathcal{V}$. En d'autres termes, on souhaite par cette opération travailler véritablement sur des CLA-k-$\{f_i\}$. Si dans la suite nous continuerons à parler de configurations de type V_2^x ou V_3^x, il est important de noter que ces dernières diffèreront sensiblement des configurations précédentes. Il s'agira en fait de ces modèles de géométrie, intersectés avec le masque précédemment décrit. Il en résultera notamment que par exemple pour une géométrie de type V_2^x, les points non contraints ne seront plus (comme dans Seaperc et Krakatao) groupés autours de la cible, mais étendus et distants de 3 km. Une telle pratique est discutable, puisqu'elle revient à effectuer un choix arbitraire de points à cadence régulière et donc au mépris de tous les problèmes de recouvrement spectral que cela peut entraîner. Nous ne la posons pas en exemple, mais comme un "tour de main" pour mener à bien l'analyse de ce PTP. Cette sélection arbitraire existait d'ailleurs déjà dans des géométries du type V_3^x puisque celles-ci équivalent à une sélection arbitraire des points frontière.

F.2 Inférence de $\mathcal{V}$, "k" et $\{f_i\}$

Nous expérimenterons successivement ces nouvelles configurations géométriques sur les zones globale et locale . Les fréquences possibles seront prises parmi l'ensemble $\{24^h,12^h\}$, bien qu'on subodore qu'elles ne soient pas significatives pour Mohns. Les exposants "global" et "local" seront utilisés pour étiqueter les ensembles de paramètres inférés en fonction de la zone concernée.

Inférence de $\{ k , \{fi\} \}^{global}$ conditionnellement à $\mathcal{V}_2^x$:

La décomposition de la route est celle des véritables pseudo-profils lâches de longueur 400 minutes, la population de $\mathcal{V}$ a été fixée à Npoint = 20 éléments dont 7 qui doivent appartenir à 7 profils différents (Ncode = 7). La configuration résultant de ces contraintes est donc celle du type $\mathcal{V}_2^x$ avec (cas particulier), une résolution de $\simeq$ 3 km le long des profils.

Conditionnellement à cette géométrie appliquée sur la zone globale, les valeurs $\emptyset$, 1 ou 2 ont été testées pour le degré k, et ceci avec ou sans la période $T = 24^h$ comme modélisation de D(t).

Ceci donne 6 ensembles de paramètres testés sur 180 voisinages. L'ensemble des résultats relatifs à cette première approche globale est synthétisé dans la figure f.1 que nous allons maintenant examiner.

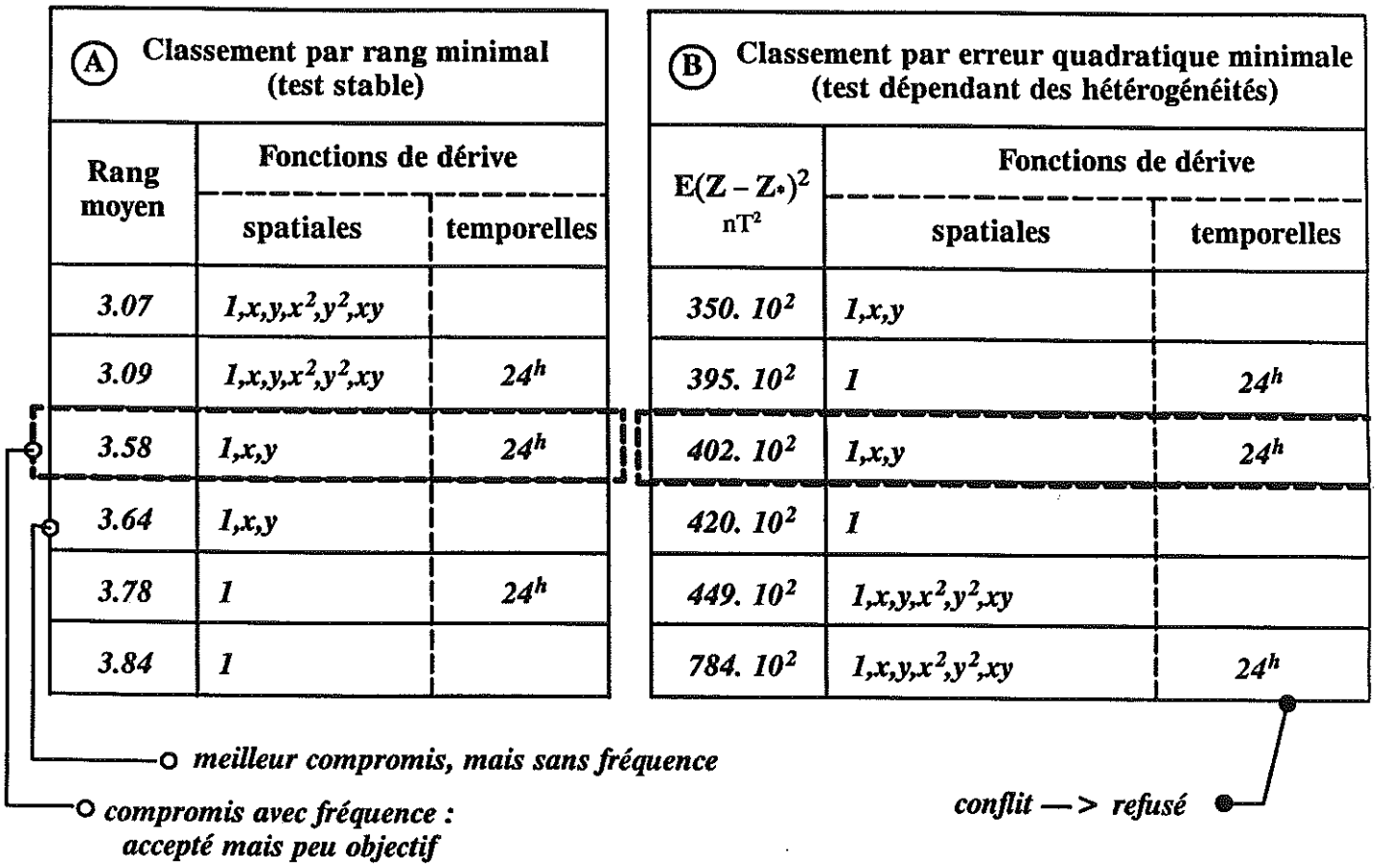


Figure f.1 : MOHNS – Magnétisme, détection de fréquences temporelles (zone globale, voisinage du type $\mathcal{V}_2^x$)

On note en premier lieu l'échelle considérable des écarts quadratiques qui s'échelonnent de 350.10^2 nT² à 784.10^2 nT², ce qui est important comparativement au domaine de variation de z(x,t) .

La comparaison des critères A (rang moyen) et B (distance quadratique) appelle 2 remarques essentielles :

- La fréquence temporelle n'est optimale pour aucun des 2 critères.
- En ce qui concerne les fonctions spatiales, il y a un conflit entre les critères A et B, puisque l'ensemble { 1, x, y } optimal pour B est loin de l'être pour A. Et réciproquement, l'ensemble { 1, x, y, x^2, y^2, xy} qui est optimal pour A, arrive en cinquième position pour B.

Il résulte de ce conflit (presque total) que ce modèle devrait être abandonné et ceci d'autant plus que le meilleur compromis, à savoir l'ensemble {1, x, y} ne contient pas la période 24^h, et rendrait donc impossible l'estimation d'une dérive temporelle dans le chapitre 4 qui utilise ces modèles. Nous choisissons donc "a priori" l'ensemble suivant de paramètres:

$$\left\{ \mathcal{V}_2^x \ ; \ k = 1 \ ; \ T = 24^h \right\}^{\text{global}}_{\text{prior}} .$$

Inférence de $\{ k , \{f_i\} \}^{\text{local}}$ conditionnellement à $\mathcal{V}_2^x$:

La même géométrie que précédemment est utilisée, et les mêmes ensembles de paramètres sont testés. La différence est qu'ici l'analyse porte sur la zone locale seulement.

Ceci donne 6 ensembles de paramètres testés sur 100 voisinages seulement, à cause de la réduction du champ d'analyse. L'ensemble des résultats relatifs à cette approche locale est synthétisé dans la figure f.2 que nous allons maintenant examiner.

Ⓐ Classement par rang minimal (test stable)		
Rang moyen	**Fonctions de dérive**	
	spatiales	**temporelles**
2.88	*1,x,y,x²,y²,xy*	
2.96	*1,x,y,x²,y²,xy*	*24h*
3.12	*1,x,y*	*24h*
3.27	*1,x,y*	
4.38	*1*	*24h*
4.39	*1*	

Ⓑ Classement par erreur quadratique minimale (test dépendant des hétérogénéités)		
$E(Z - Z_*)^2$ nT^2	**Fonctions de dérive**	
	spatiales	**temporelles**
138. 10²	*1,x,y,x²,y²,xy*	
142. 10²	*1,x,y*	*24h*
148. 10²	*1,x,y,x²,y²,xy*	*24h*
149. 10²	*1,x,y*	
266. 10³	*1*	*24h*
330. 10²	*1*	

bon compromis avec fréquenc —> accepté

Figure f.2 : Détection de fréquences temporelle (zone locale, voisinage du type $\mathcal{V}_2^x$)

On note tout d'abord que si les écarts quadratiques sont toujours de l'ordre de 10^4, ils s'échelonnent cette fois-ci sur l'intervalle $[138.10^2\ nT^2, 330.10^2\ nT^2]$, ce qui est nettement meilleur que lors de l'analyse sur tout le champ. Il s'agit de savoir si un tel gain de précision est lié à l'hétérogénéité de l'échantillonnage ou à la variable elle-même. Les diamètres moyens de $\mathcal{V}$ dans l'un et l'autre cas, permettent de lever le voile : il sont en effet de 41 km pour la zone globale et de 17 km pour la zone locale. Ce qui diffère dans ces deux zones, c'est bien l'interdistance des profils qui est considérable hors de la zone locale. Il en résulte que les contraintes du type $\mathcal{V}_2^x$ (7 points sur 7 profils) forcent la machine à aller chercher des points considérablement éloignés, et ce sont ces cas de figure qui "tirent" le diamètre moyen vers la valeur $\mathcal{D}^x$ = 41 km et expliqueraient les différences de variance entre ces deux zones.

Le choix du modèle, quant à lui, ne présente pas de conflit comme précédemment, puisque à "l'unanimité" (entre A et B) c'est le modèle $\{1, x, y, x^2, y^2, xy\}$ qui est le meilleur dans les deux cas. Ceci montre que c'était l'hétérogénéité de l'échantillonnage qui engendrait un conflit dans l'inférence globale.

On constate qu'une fois de plus, la fréquence temporelle ne fait pas partie de l'ensemble optimal. Son adoption rend donc ce paramètre "a priori". L'ensemble suivant sera adopté :

$$\left\{ \mathcal{V}_2^x \ ;\ k = 2 \ ;\ T_{prior} = 24^h \right\}^{local}.$$

Inférence de $\{ k , \{fi\} \}^{global}$ conditionnellement à $\mathcal{V}_3^x$:

La décomposition de la route est celle de pseudo-profils courts (Δcode = 1 heure) avec une population de $\mathcal{V}$ portée à Npoint = 20 points qui doivent appartenir à 20 pseudo-profils différents (Ncode = 20). La configuration résultante est celle du type V_3^x avec (cas particulier), une résolution minimale de $\simeq$ 3 km le long des profils. Conditionnellement à cette géométrie appliquée sur la zone globale, les valeurs $\emptyset$, 1 ou 2 ont été testées pour le degré k, et ceci avec ou sans les périodes 24^h et 12^h pour la modélisation de D(t). Ceci donne 9 ensembles de paramètres testés sur 210 voisinages. Les résultats sont présentés dans la figure f.3.

(A) Classement par rang minimal (test stable)

Rang moyen	Fonctions de dérive	
	spatiales	temporelles
3.20	*1,x,y,x²,y²,xy*	
3.21	*1,x,y,x²,y²,xy*	*24h, 12h*
3.31	*1,x,y,x²,y²,xy*	*24h*
3.42	*1,x,y*	*24h, 12h*
3.44	*1,x,y*	*24h*
3.55	*1*	*24h*
3.60	*1,x,y*	
3.67	*1*	*24h, 12h*
3.91	*1*	

adopté mais arbitraire

(B) Classement par erreur quadratique minimale (test dépendant des hétérogénéités)

$E(Z - Z_*)^2$ nT^2	Fonctions de dérive	
	spatiales	temporelles
486. 10²	*1,x,y*	
637. 10²	*1,x,y*	*24h*
650. 10²	*1*	*24h, 12h*
655. 10²	*1*	*24h*
676. 10²	*1*	
768. 10²	*1,x,y*	*24h, 12h*
104. 10³	*1,x,y,x²,y²,xy*	*24h, 12h*
131. 10³	*1,x,y,x²,y²,xy*	
155. 10³	*1,x,y,x²,y²,xy*	*24h*

Figure f.3 : Détection de fréquences temporelles (zone globale, voisinage du type $\mathcal{V}_3^x$)

On constate tout d'abord que l'échelle des variances s'étend sur l'intervalle [486.10^2 nT^2, 155.10^3 nT^2] ce qui constitue une nette régression par rapport aux cas précédents, local et global confondus. Ce fait est sans doute corrélé à l'augmentation du diamètre moyen de $\mathcal{V}$ qui est passé à 65 km .

De plus, on constate qu'une fois encore, il y a conflit entre les critères A et B, et que les fréquences temporelles n'ont pas pu être détectées. Les paramètres suivants sont donc intégralement "a priori" :

$$\left\{ \mathcal{V}_3^x \ ; \ k = 1 \ ; \ T = 24^h \right\}_{prior}^{global} .$$

Inférence de { k , {fi} } ^{local} conditionnellement à φ_3^x :

Les mêmes paramètres de voisinage que précédemment sont utilisés, de même que les mêmes ensembles de paramètres sont testés.

La différence est qu'ici l'analyse porte sur la zone locale seulement. Ceci donne 9 ensembles de paramètres testés sur 90 voisinages seulement, à cause de la réduction du champ d'analyse. L'ensemble des résultats relatifs à cette approche locale est synthétisé dans la figure f.4 que nous allons maintenant examiner.

(A) Classement par rang minimal (test stable)

Rang moyen	Fonctions de dérive	
	spatiales	temporelles
2.88	*1,x,y,x²,y²,xy*	
2.93	*1,x,y,x²,y²,xy*	*24h, 12h*
3.02	*1,x,y,x²,y²,xy*	*24h*
3.35	*1,x,y*	*24h, 12h*
3.35	*1,x,y*	*24h*
3.45	*1,x,y*	
3.96	*1*	
4.10	*1*	*24h, 12h*
4.34	*1*	*24h*

compromis adopté car contenant des fréquences

(B) Classement par erreur quadratique minimale (test dépendant des hétérogénéités)

$E(Z - Z_*)^2$ / nT^2	Fonctions de dérive	
	spatiales	temporelles
517. 10²	*1,x,y,x²,y²,xy*	
575. 10²	*1,x,y,x²,y²,xy*	*24h, 12h*
611. 10²	*1,x,y,x²,y²,xy*	*24h*
781. 10²	*1,x,y*	
803. 10²	*1,x,y*	*24h, 12h*
869. 10²	*1,x,y*	*24h*
124. 10³	*1*	
144. 10³	*1*	*24h, 12h*
147. 10³	*1*	*24h*

meilleur modèle, mais refusé car ne contenant pas de fréquences

Figure f.4 : Détection de fréquences temporelles (zone locale, voisinage du type φ_3^x)

Comme précédemment, le passage du global au local a permis de lever les conflits existants entre les deux critères, puisque c'est k = 2 (mais sans fréquence) qui arrive en meilleure position dans les deux cas. Nous sommes donc contraint d'adopter l'ensemble suivant de paramètres :

$$\left\{ \varphi_3^x \ ; \ k = 2 \ ; \ T_{prior} = 24^h, 12^h \right\}^{global},$$

bien que cet ensemble ne représente pas l'optimalité. On notera cependant, que parmi les ensembles contenant des fréquences temporelles, il y a accord entre les deux critères.

On remarquera que cette fois-ci, la fréquence $T_2 = 12^h$ a été adoptée.

Bilan :
Les 4 ensembles suivants de paramètres sont désormais disponibles:

- $\left\{ \mathcal{V}_2^x \; ; \; k = 1 \; ; \; T = 24^h \right\}_{\text{prior}}^{\text{global}}$,
- $\left\{ \mathcal{V}_2^x \; ; \; k = 2 \; ; \; T_{\text{prior}} = 24^h \right\}^{\text{local}}$,
- $\left\{ \mathcal{V}_3^x \; ; \; k = 1 \; ; \; T = 24^h \right\}_{\text{prior}}^{\text{global}}$,
- $\left\{ \mathcal{V}_3^x \; ; \; k = 2 \; ; \; T_{\text{prior}} = 24^h, 12^h \right\}^{\text{global}}$,

Leur trait commun est essentiellement l' "a priori" des fréquences temporelles qui ont été imposées dans les 4 cas puisque leur détection n'a pu être possible. Ce résultat confirme ce que nous savions de z(x,t) par l'intermédiaire de la série temporelle Mohns, à savoir que sous cette latitude, D(t) n'est pas périodique. On peut donc s'estimer satisfait que nos inférences confirment ce fait expérimental.

Puisque Mohns n'est pas un PTP, il aurait fallu renoncer à modéliser D(t) par des fonctions trigonométriques. Un modèle de FAI-k spatio-temporelle s'avérerait sans doute plus adapté à ce phénomène. Nous ne le ferons pas car notre objectif est ici de voir ce qu'il advient lorsqu'on force les fonctions trigonométriques à s'ajuster sur des fluctuations non périodiques (en tendance). La suite de ce cas d'étude ainsi que les estimations du chapitre 4 qui en résulteront, ne représentent donc qu'une curiosité scientifique et certainement pas une modélisation utile (ni ultime).

Si on compare les deux ensembles qui ne sont pas intégralement "a priori" , à savoir ceux portant sur la zone locale, on constate que dans les 2 cas, le degré inféré est k = 2, et ceci pour des géométries très différentes. Ce fait confirme une fois de plus la robustesse (section 2.3) de ce paramètre.

Examinons maintenant la figure f.5 qui résume les caractéristiques de ces 4 ensembles. Deux points essentiels se dégagent de son analyse :

- à $\mathcal{V}$ fixé, le passage du global au local s'accompagne toujours d'une diminution de l'erreur quadratique. En effet, elle passe de 402.10^2 nT^2 à 148.10^2 nT^2 pour $\mathcal{V} = \mathcal{V}_2^x$ et de 768.10^2 nT^2 à 575.10^2 nT^2 dans l'autre cas.
- les variances croissent avec le diamètre $\mathcal{D}$, ceci étant vrai à l'exception du modèle quatre, sans doute instable du fait de l'introduction de $T_2 = 12^h$ dans le système.

Si les rangs moyens suivent la hiérarchie donnée par les variances , on rappellera que ceux-ci ne sont pas comparables entre les différents modèles. On peut cependant conjecturer qu'un rang 3.44, par rapport à 9 cas possibles (modèle trois), est certainement plus favorable qu'un rang de 3.58 par rapport à 6 cas possibles (modèle un) , mais ce genre d'analyse est discutable.

Comme de toute façon, cette analyse est biaisée par le fait qu'on a forcé ces quatre modèles à contenir des fréquences temporelles, nous préférons ne pas nous étendre plus sur ces résultats.

	Modèle un	Modèle deux	Modèle trois	Modèle quatre
Zone	**Globale**	**Locale**	**Globale**	**Locale**
Type de $\mathcal{V}$	$\mathcal{V}_2^x$	$\mathcal{V}_2^x$	$\mathcal{V}_3^x$	$\mathcal{V}_3^x$
$\mathcal{D}^x$ moyen	**41 km**	**17 km**	**65 km**	**31 km**
"k" choisi	**1**	**2**	**1**	**2**
"T_i" choisies	**$T_1 = 24^h$**	**$T_1 = 24^h$**	**$T_1 = 24^h$**	**$T_1 = 24^h$** **$T_2 = 12^h$**
Rang moyen	**3.58 (/6)**	**2.96 (/6)**	**3.44 (/9)**	**2.93 (/9)**
Distance quadratique	**$402.\ 10^2\ nT^2$**	**$148.\ 10^2\ nT^2$**	**$637.\ 10^2\ nT^2$**	**$575.\ 10^2\ nT^2$**

Figure f.5 : Synthèse sur la détection de fréquences temporelles

F.3 Inférence et validation de K(h)

Il reste maintenant à associer des covariances généralisées à chacun des quatre modèles précédents. On rappelle que si on continuera à parler de voisinages de type $\mathcal{V}_2^x$ ou $\mathcal{V}_3^x$, ce ne seront pas exactement ces géométries qui seront impliquées, mais celles-ci appliquées à une présélection arbitraire de 1 donnée sur 10.

Inférence de K(h) cond. à $\left[\mathcal{V}_2^x\ ;\ k = 1\ ;\ T = 24^h\right]_{prior}^{global}$ **:**

L'ensemble des paramètres $\left\{\ \mathcal{V}_2^x\ ;\ k = 2\ ;\ T_{prior} = 24^h\ \right\}^{local}$ étant fixé, parmi tous les ensembles de covariance généralisée à notre disposition, les trois modèles suivants ont finalement été retenus selon leurs critères de Jack-knife :

- $K_1^1(h) = a_1\ |h|$
- $K_2^1(h) = a'_1\ |h| + a'_3|h|^3$
- $K_3^1(h) = a''_1\ |h| + a''_2|h|^2\ Log(|h|)$

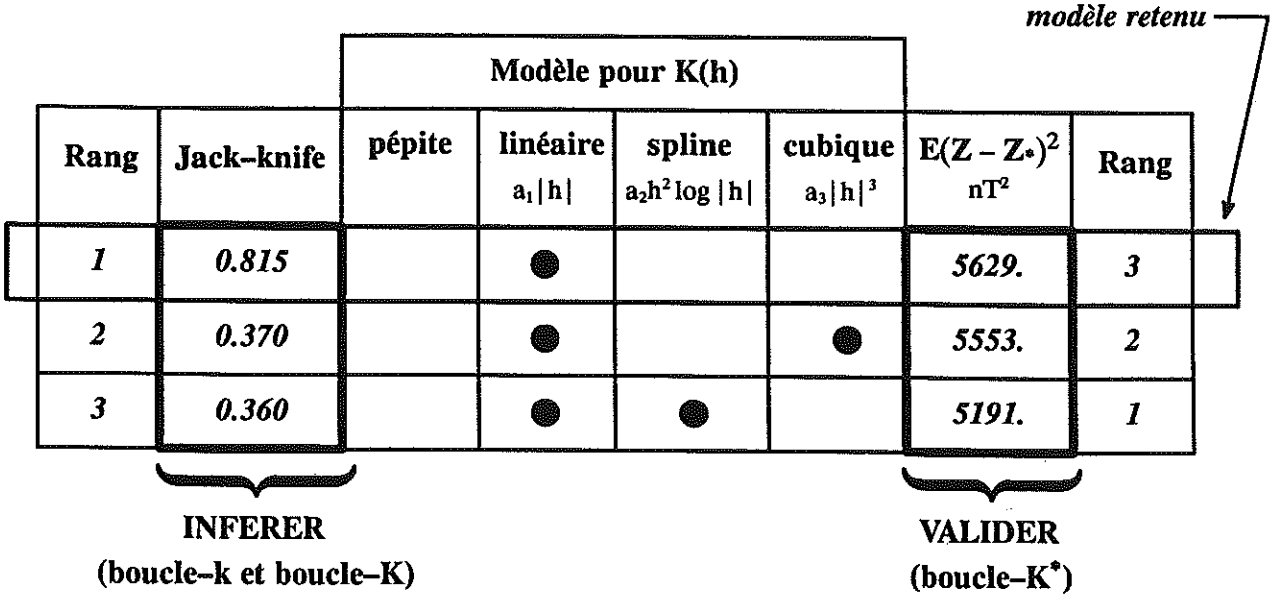


Figure f.6 : Inférence de la covariance généralisée K (h)
(zone globale, conditionnellement au premier ensemble de paramètres)

La figure f.6 contient la classification de ces trois modèles, par leurs Jack-knife, et leurs variances expérimentales. On y note que le premier modèle est de loin le meilleur par son Jack-knife proche de 1. Ce résultat est incohérent avec la classification donnée par le critère des variances .

A propos de ces dernières, on note en premier lieu qu'elles sont sans commune mesure avec celles des précédents cas d'études, et ceci sans doute à cause de la présélection arbitraire d'une donnée sur 10.

On note ensuite que si le premier ensemble est optimal par son Jack-knife, il se trouve en dernière position selon le critère de variance, de sorte que le conflit est total entre l'inférence et la validation. Mais les fluctuations relatives des variances sont faibles par rapport à celles des Jack-knife, de sorte que le premier modèle est tout à fait acceptable.

En conséquence, conditionnellement à l'ensemble de paramètres $\left\{ \mathcal{V}_2^x \ ; \ k = 1 \ ; \ T = 24^h \right\}_{\text{prior}}^{\text{global}}$, c'est le modèle suivant de CG. qui sera adopté :

$K^1(h) = a_1 |h|$ avec $a_1 = -4990$

Inférence de K(h) cond. à $\left\{\mathcal{V}_2^x \ ; \ k = 2 \ ; \ T_{\text{prior}} = 24^h\right\}^{\text{local}}$:

L'ensemble $\left\{ \mathcal{V}_2^x \ ; \ k = 2 \ ; \ T_{\text{prior}} = 24^h \right\}^{\text{local}}$ étant fixé, la présélection (par Jack-knife) des différents modèles K(h) possibles, a permis de nous focaliser sur les trois modèles suivants :

- $K_1^2(h) = a_\beta \ \text{pépite}(h) + a_3 \ |h|^3$
- $K_2^2(h) = a'_\beta \ \text{pépite}(h) + a'_2 \ |h|^2 \ \text{Log}(|h|)$
- $K_3^2(h) = a''_1 \ |h| + a''_3 |h|^3$

modèle retenu

		Modèle pour K(h)					
Rang	**Jack-knife**	**pépite**	**linéaire** $a_1\|h\|$	**spline** $a_2h^2\log\|h\|$	**cubique** $a_3\|h\|^3$	**$E(Z-Z^*)^2$** nT^2	**Rang**
1	*0.983*	●			●	*14830*	*2*
2	*1.026*	●		●		*14760*	*1*
3	*1.080*		●		●	*15480*	*3*

INFERER (boucle-k et boucle-K)

VALIDER (boucle-K^*)

Figure f.7 : Inférence de la covariance généralisée K (h)
(zone locale, conditionnellement au deuxième ensemble de paramètres)

Ce qui frappe d'emblée sur la figure f.7, ce sont les valeurs extrêmes des variances expérimentales qui, de l'ordre de $\simeq 5.10^3\ nT^2$ (modèle global) sont passées à $\simeq 15.10^3$ (modèle local).

Ce fait est d'autant plus étonnant qu'il s'accompagne d'une amélioration sensible des Jack-knife qui, de l'ordre de 0.8 (zone globale) sont passés à $\simeq 0.98$ (zone locale). Il y a donc une inquiétante contradiction avec les expériences précédentes (moindre carré) où le passage du global au local s'accompagnait d'un gain sensible de variance. Mais si on y regarde de plus près, en comparant notamment la figure f.5 aux figures f.6 et f.7, on constate que les valeurs en jeu pour cette zone locale, sont de l'ordre de celles en jeu dans une estimation par moindre carré : ($148.10^2\ nT^2$ pour l'inférence de k, 148.10^2 pour la validation de K(h)). C'est donc la valeur précédente de 5629 nT^2 (zone globale) qui était anormalement faible et on pourrait l'attribuer (pourquoi pas) à l'avantage d'utiliser l'estimateur par FAI-k au lieu d'une opération de moindre carré. Mais il est à craindre que ce soient plutôt nos fréquences temporelles, inadaptées à ce cas d'étude, qui introduisent des artefacts.

Comme précédemment, il y a conflit entre les deux critères. Les modèles inférés dans ces circonstances sont donc arbitraires.

Conditionnellement à l'ensemble de paramètres $\left\{ \mathcal{V}_2^x\ ;\ k=2\ ;\ T_{prior}=24^h \right\}^{local}$, c'est le modèle (arbitraire) suivant de CG. qui sera adopté :

$K^2(h) = a_\emptyset$ pépite$(h) + a_3\ |h|^3$ avec $a_\emptyset = 1\,151.$ et $a_3 = 179.9,$

Inférence de K(h) cond. à $\left[\mathcal{V}_3^x\ ;\ k=1\ ;\ T=24^h\right]^{global}_{prior}$:

L'ensemble $\left\{ \mathcal{V}_3^x\ ;\ k=1\ ;\ T=24^h \right\}^{global}_{prior}$ étant fixé, les 3 modèles suivants ont finalement été retenus :

- $K_1^3(h) = a_1\ |h|$
- $K_2^3(h) = a'_1\ |h| + a'_2\ |h|^2\ \mathrm{Log}(|h|)$
- $K_3^3(h) = a''_\emptyset$ pépite$(h) + a''_3\ |h|^3$

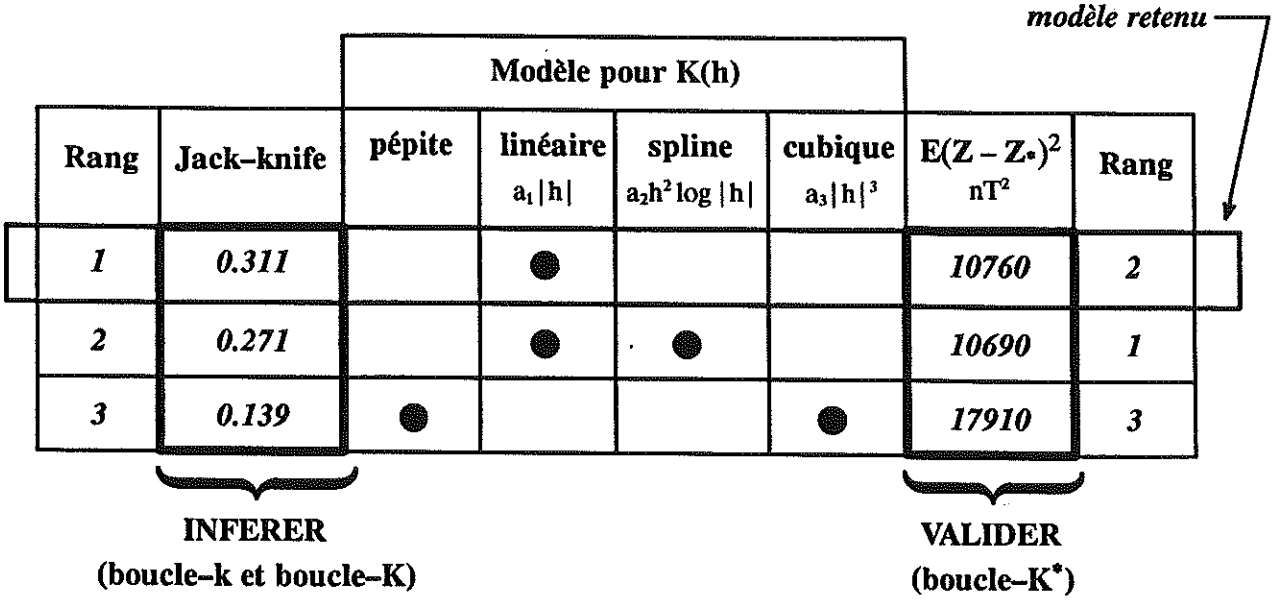


Figure f.8 : Inférence de la covariance généralisée K (h)
(zone globale, conditionnellement au troisième ensemble de paramètres)

La figure f.8 qui synthétise ces résultats montre des Jack-knife très faibles. De plus, on constate qu'il y a une fois de plus conflit entre les deux critères. Bref, nous sommes toujours dans la situation difficile de modèles manifestement inadaptés à nos données et c'est d'une manière intégralement arbitraire que nous adopterons le modèle suivant de CG :

$K^3(h) = a_1 |h|$ avec $a_1 = -14\,870$

Inférence de K(h)cond à $\left\{\mathcal{V}_3^x ; k = 2 ; T_{prior} = 24^h, 12^h\right\}^{local}$ **:**

L'ensemble $\left\{ \mathcal{V}_3^x ; k = 2 ; T_{prior} = 24^h, 12^h \right\}^{local}$ étant fixé, les tests préliminaires ont conduit aux 3 modèles suivants de CG :

- $K_1^4(h) = a_1 |h|$
- $K_2^4(h) = a'_1 |h| + a'_3|h|^3$
- $K_3^4(h) = a''_1 |h| + a''_2 |h|^2 \text{ Log}(|h|)$

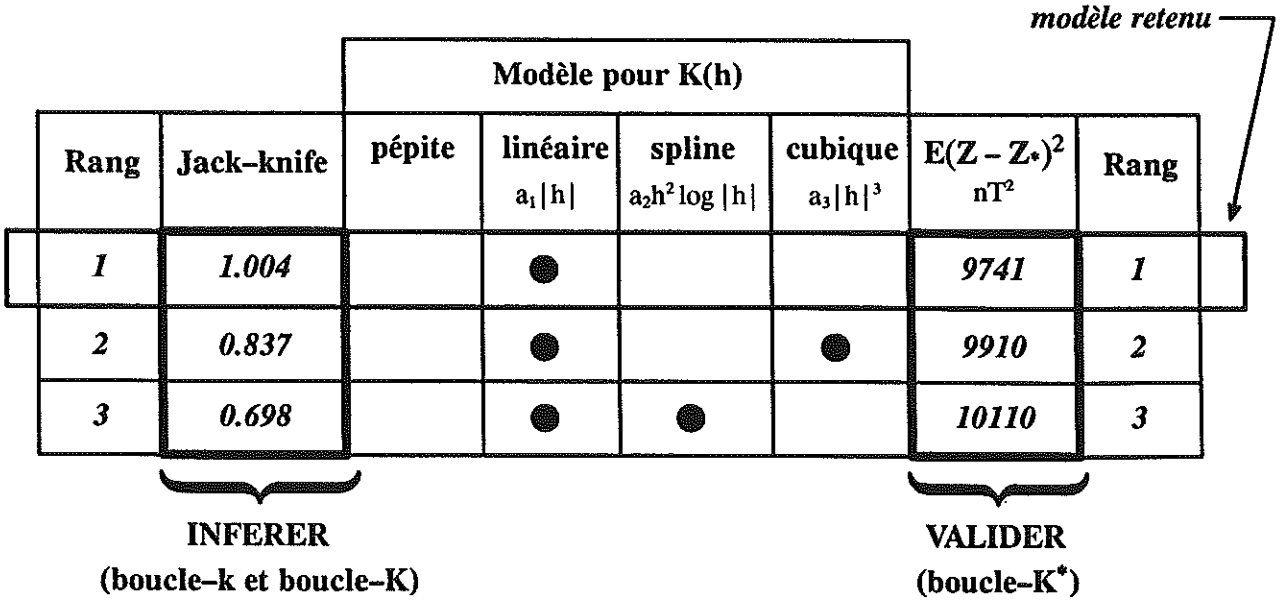


Figure f.9 : Inférence de la covariance généralisée K (h)
(zone locale, conditionnellement au quatrième ensemble de paramètres)

L'analyse de la figure f.9 permet de constater qu'il y a cohérence totale entre les critères de Jack-knife et de variance. Cette cohérence s'accompagne d'un Jack-knife particulièrement favorable (1.004) et d'une diminution sensible de la variance expérimentale, par rapport au cas précédent. On adoptera donc le modèle suivant :

$K^4_{mohns}(h) = a_1 |h|$ avec $a_1 = -111\,90$.

F.4 Conclusions

Les analyses ont permis d'aboutir à 4 ensembles de paramètres:

- $\left\{ \mathcal{V}_2^x ;\ k=1 ;\ T=24^h ;\ K^1(h) \right\}^{\text{global}}_{\text{prior}}$

avec $K^1(h) = a_1 |h|$

- $\left\{ \mathcal{V}_2^x ;\ k=2 ;\ T_{prior}=24^h ;\ K^2(h) = K^{prior}(h) \right\}^{\text{local}}$

avec $K^2(h) = a_\emptyset$ pépite$(h) + a_3 |h|^3$

- $\left\{ \mathcal{V}_3^x ;\ k=1 ;\ T=24^h ;\ K^3(h) = K^{prior}(h) \right\}^{\text{global}}_{\text{prior}}$

avec $K^3(h) = a_1 |h|$

- $\left\{ \mathcal{V}_3^x ;\ k=2 ;\ T_{prior}=24^h, 12^h ;\ K^4(h) \right\}^{\text{local}}$

avec $K^4(h) = a_1 |h|$

	Modèle un (global, V_2^x)	Modèle deux (local, V_2^x)	Modèle trois (global, V_3^x)	Modèle quatre (local, V_3^x)
K(h)	-4990 |h|	-1151 pépite|h| + -179.9 |h|^3	-14870 |h|	-11190 |h|
Jack-knife	3 0.815	2 0.983	4 0.311	1 1.004
Variance expérimentale	1 5629 nT^2	4 14830 nT^2	3 10760 nT^2	2 9741 nT^2
	4	6	7	3

plus mauvais modèle

meilleur modèle

Figure f.10 : Synthèse sur l'inférence et la validation de K(h)

Si l'on examine la figure f.10 qui résume les résultats obtenus pour ces quatre modèles, on constate que le passage du global au local s'accompagne toujours d'une amélioration sensible du Jack-knife. Or, le Jack-knife est, on le rappelle, un test de cohérence entre une variance théorique et une variance expérimentale. Ce premier résultat signifie donc que la zone dite locale est plus adaptée à un modèle unique que la zone dite globale ; ce qui est une évidence, certes, mais n'en constitue pas moins une satisfaction dans cette fastidieuse analyse structurale. Par contre, ce même tableau montre qu'en terme de variance expérimentale, il n'est pas de même, puisque, selon la géométrie de γ, le passage du global au local s'accompagne d'une augmentation de variance expérimentale (pour γ_2^x) ou d'une diminution (pour γ_3^x). Ceci semble signifier :

- Conditionnellement à γ_2^x, le système d'analyse est beaucoup plus instable sur la zone locale.

- Conditionnellement à γ_3^x, le système d'analyse est instable sur le complémentaire de la zone locale dans la zone globale (ie les extrémités du champ informées par des profils peu resserrés.

Si maintenant on compare entre eux les Jack-knife et les variances expérimentales issues de la "boucle-K*", on constate que :

- Pour les modèles 2 et 4, de bons Jack-knife (car proches de 1) s'accompagnent de variances expérimentales importantes.

- Pour le modèle 1, le Jack-knife (non optimal) s'accompagne d'une variance minimale.

Or, il est bon de rappeler qu'un Jack-knife est approximativement le rapport d'une erreur quadratique (estimation par moindre carré) à la variance théorique d'une CLA-k-$\{f_i\}$ établie à l'aide de la covariance testée. De même, rappelons que les variances expérimentales de la "boucle-K*" sont celles d'écarts entre une mesure et son estimation obtenue par le modèle testé. Si donc il y a conflit entre Jack-knife et variance expérimentale, cela semble signifier que pour ce jeu de données, un modèle sera d'autant meilleur qu'il s'éloignera d'une opération de moindre carré.

De même, ce conflit peut signifier que l'ensemble des modèles disponibles de CG est insuffisant pour appréhender ce PTP. Ces conjectures rejoignent les conclusions précédentes à savoir qu'ici, c'est un modèle de FAI-k spatio-temporelle qu'il faudrait inférer, avec notamment une CG K(h,t) de la forme :

$$K(h, t) = K_1(h) + \sigma_D(t)$$

Bien que ces modèles servent à des tentatives de filtrage de D(t), nous savons d'ores et déjà que modéliser D(t) par une série de Fourier est un choix inadapté au jeu de données Mohns. La conclusion de ce cas d'étude peut donc se résumer à cette seule phrase: Mohns n'est pas un PTP.

Nomenclature des notations, abréviations et mots-clefs

Notation, abréviation	Mot-clef, définition	page
A - B		
AK	**Analyse Krigeante**	200
AKT	**Analyse Krigeante Trigonométrique**	200
	Bochner (théorème de ...)	29
BLUEPACK	**Best Linear Unbiaised Estimator Package**	187
Boucle-k	**Boucle d'inférence** du degré k de la Fonction Aléatoire Intrinsèque	189
Boucle-K	**Boucle d'inférence** de la Covariance Généralisée K(h)	191
Boucle-K*	**Boucle de validation** de la Covariance Généralisée K(h)	193
Boucle-k-$\{f_i\}$	**Boucle d'inférence** du degré k et des fréquences f_i de la Fonction Aléatoire Intrinsèque	71
C - D		
CLA-k	**Combinaison Linéaire** Autorisée à l'ordre k	186
CLA-k-$\{f_i\}$	**Combinaison Linéaire** Autorisée à l'ordre des polynômes de degré k et des fréquences f_i .	70
	Composante périodique d'un Phénomène à Tendance Périodique	27
	Composante non périodique d'un Phénomène à Tendance Périodique	27
C_D	**Covariance** stationnaire de D(.)	39
CG	**Covariance** Généralisée (abréviation)	186
k	**Degré** polynomial maximal de la Fonction Aléatoire Intrinsèque	185
D_{TFD}	**Diamètre** de la fenêtre d'analyse de la TFD	182
$\mathfrak{D}$	**Diamètre** du voisinage d'analyse et d'estimation	72
	Diffusion de l'énergie	181
E - F		
	Effet de route	83
ξ	**Energie** d'une fonction	16
	Espace-période du PTP	16
Icode	**Etiquette** d'un pseudo-profil	69
Ivois	**Etiquette** d'un voisinage test	189
FFT	**Fast Fourier Transformer**	179
FA	**Fonction Aléatoire**	183
FAI-k	**Fonction Aléatoire** Intrinsèque d'ordre k	185
FAI-G	**Fonction Aléatoire** Intrinsèque Généralisée	187
FAI-$\mathfrak{F}$	**Fonction Aléatoire** Intrinsèque Généralisée à l'ordre de l'espace $\mathfrak{F}$	187
FAST	**Fonction Aléatoire** Stationnaire d'ordre 2	183

Notation, abréviation	Mot-clef, définition	page
G - L		
	Glissant (voisinage ...)	187
	Instabilité (d'un krigeage)	198
Q*	**Jack-knife**	192
K(h)	**Covariance** Généralisée (symbole)	185
	Krigeage	195
KT	**Krigeage** Trigonométrique	199
Δcode	**Longueur** du pseudo-profil	69
M - O		
	Modulation	17
Npoint	**Nombre** de points du voisinage d'analyse et d'estimation	187
Nvois	**Nombre** de voisinages tests	189
Ncode	**Nombre** de pseudo-profils	75
P - Q		
	Palier d'un variogramme	184
$\mathcal{P}$	**Pas** d'échantillonnage du voisinage d'analyse et d'estimation	72
P_{TFD}	**Pas** d'échantillonnage de la fenêtre d'analyse de la TFD	182
$\gamma_{pép}(h)$	**Pépitique** (modèle de variogramme ...)	184
pépite(h)	Indicatrice de 0	184
T_e	**Période** d'échantillonnage d'une fonction	179
$\gamma_{Pério}(h)$	**Périodique** (modèle de variogramme ...)	184
	Périodogramme	32
PTP	**Phénomène** à Tendance périodique	11
	Positivité conditionnelle	186
Polynomiale	**Modèle** de Covariance Généralisée ...	186
$\{\lambda_\alpha\}_\alpha$	**Pondérateurs** de Krigeage	195
	Portée finie	184
	Portée infinie	184
	Portée pratique	184
	Positivité conditionnelle stricte	72
Prior	**A priori** ($k_{...}$, $K_{...}$, $T_{...}$)	189
	Pseudo-profil	69
$\gamma_{Puis}(h)$	**Puissance** (modèle de variogramme ...)	184
R - S		
$[\mathcal{V}(x_i)]_{R_1}$	**Rangs** impairs (éléments du voisinage étendu, de ...)	189
$[\mathcal{V}(x_i)]_{R_2}$	**Rangs** pairs (éléments du voisinage étendu, de ...)	189

Notation, abréviation	Mot-clef, définition	page
R – S ...suite		
$rep_T\ \{ z(t, t_1, T) \}$	**Répétition** périodique de période T de la restriction de z(t) à $[t_1 , t_1 + T]$	180
	Robuste (paramètre ...)	57
$S\mathcal{F}$	**Série** de Fourier	70
$\gamma_{sphé}(h)$	**Sphérique** (modèle de variogramme ...)	184
$S_{kprior}(ivois)$	**Variance** expérimentale de la zone test "ivois".	189
	Spline (modèle de Covariance Généralisée ...)	186
$\sigma_D(.)$	**Structure** à palier du variogramme de la fonction D(.)	39
T – U		
$T\mathcal{F}$	**Transformée** Intégrale de Fourier (considérée comme opérateur)	29
TFD	**Transformée** de Fourier Discrète	179
TFR	**Transformée** de Fourier Rapide	179
T – U ...suite		
$\gamma_{Trou}(h)$	**Trou** (modèle de variogramme à effet de ...)	184
	Unique (voisinage ...)	187
V – Z		
VR	**Variable** Régionalisée	15
$\gamma(h)$	**Variogramme**	183
$\gamma^*(h)$	**Variogramme** expérimental	185
$\mathcal{V}$	**Voisinage** d'analyse et d'estimation	187
$\mathcal{V}_1^x$	**Voisinage** de géométrie "spatiale aveugle"	75
$\mathcal{V}_2^x$	**Voisinage** de géométrie "spatiale anisotrope "	75
$\mathcal{V}_3^x$	**Voisinage** de géométrie "spatiale isotrope sautante"	76
$\mathcal{V}_4^{x\ t}$	**Voisinage** de géométrie "spatio-temporelle"	77
$[\mathcal{V}(x_i)]_{étendu}$	**Voisinage** étendu spatialement	189

Bibliographie et références

[GΦ : cité pour son intérêt géophysique]

[GST : cité pour son intérêt géostatistique]

[TdS : ouvrages de théorie du signal]

[I : institut, laboratoire]

[GΦ - 1]
- F. **GOHIN**, nov. 1987, thèse
- Analyse géostatistique des champs thermiques de la surface de la mer.
- ENSMP, centre de Géostatistique, Fontainebleau.

[GΦ - 2]
- D. W. **HANDSCHUMACHER**, 1976, Géophy. monog.
- Post-Eocene plate tectonics of the Eastern Pacifics
- Dans "The geophysics of the pacific ocean Basin and its Margin ". Edité par G.H. SUTTON, M. MANGHANI et R. MOBERLY .
  Amer. Géophys. Union, WashintonD.C., pp 177-202.

[GΦ - 3]
- F.H. **HIBBERD**, 1985
- The geomagnetic Sq variation - annual, semi-annual and solar cycle variations and ring current effects.
- J. Atoms. Terr. Phys., 47, pp 341-352

[GΦ - 4]
- Ph. **HUCHON** , J. BOURGEOIS, 10 juin 1990.
- Subduction-Inducted Fragmentation of the Nazca Plate off Peru : Mendana Fracture Zone and Trujillo Trough Revisited
- J.G.R., VOL. 95, NO B6, pp 8419-8436.

[GΦ - 5]
- Ch. **KIRSCH** , juin 1982, ref. S-120
- Etude géostatistique des données de pollution de la ville de Berlin
- ENSMP, centre de Géostatistique, Fontainebleau.

[GΦ - 6]
- X. **LE PICHON**, J. FRANCHETEAU, J. BONNIN, 1973
- Plate tectonics
- Ed Elsevier.

[GΦ - 7]
- E. K. **SMITH**, 1967
- Température - latitude sporadic E physics of Geomagnetic Phenomena.
- Ed. par S. MATSUSHITA et W.H. CAMPBELL, Academic Press, New York .

[GST - 1]
- J.P. **CHILES**, 1978, ref N-584
- L'inférence statistique automatique des FAI-k
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 2]
- A. **GALLI**, O. DUBRULE, mars 1980, ref. N-651
- Simulations (dans le plan) de Fonctions Aléatoires Intrinsèques Générales
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 3]
- A. **GALLI**, 1984, ref N-818
- Factorial Kriging Analysis : a substitute to Spectral Analysis of magnetic data
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 4]
- A.G. **JOURNEL**, 1977, monog.
- Géostatistique Minière (2 tomes)
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 5]
- Ch. **LANTUEJOUL**, août 1990, ref N-17/90/G
- Ergodicité et portée intégrale
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 6]
- G. **MATHERON**, 1970, fasc. 5
- La Théorie des Variables Régionalisées et ses applications.
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 7]
- G. **MATHERON**, oct. 1971, ref N-117
- Les Fonctions Aléatoires Intrinsèques Généralisées
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 8]
- G. **MATHERON**, oct. 1972, ref N-299
- Les Covariances Généralisées plolynomiales
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 9]
- G. **MATHERON**, juin 1974, ref N-374
- Représentations stationnaires et représentations minimales pour les FAI-k
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 10]
- G. **MATHERON**, fev. 1978, monog.
- Estimer et choisir
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 11]
- G. **MATHERON**, oct. 1979, ref N-617
- Comment translater les catastrophes ou la structure des FAI Générales
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 12]
- D. **RENARD**, avril 1983, ref. N-812
- Smoothing discontinuities when kriging with moving neighbourhood
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 13]
- S.A. **SEGURET**, sept. 1987, ref S-229/M1
- Filtrage des perturbations diurnes par dérive externe trigonométrique
- ENSMP, Centre de Géostatistique, Fontainebleau.

[GST - 14]
- S.A. **SEGURET**, Ph. HUCHON dec. 1990
- Trigonometric Kriging : a new method for removing the diurnal variation from geomagnetic data
- J.G.R., VOL. 95, NO. B13, pp 21,383-21,397

[GST - 15]
- **Service Informatique** du Centre de Géostatistique
- BLUEPACK manuel
- ENSMP, Centre de Géostatistique, Fontainebleau.

[TdS - 1]
- W.B. **DAVENPORT**, W.L. ROOT, 1958, monog.
- An introduction to the Theory of random signals and noise.
- Ed. Mc Graw-Hill, New York.

[TdS - 2]
- M. KUNT, 1980, monog.
- Traitement numérique des signaux
  Ecole polytechnique de Lausanne, volume XX
- Ed. Géorgi, Lausanne.

[TdS - 3]
- B. PICINBONO, 1977, monog.
- Eléments de Théorie du Signal
- Ed. Bordas, Paris.

[TdS - 4]
- A.J. RACK, 1938
- Effect of Space–charge and Transit time of the shot noise in diodes.
- Bell Syst. Tech. Jour., 17, pp 592–619

[TdS - 5]
- C.E. SHANNON, 1948
- The Mathematical Theory of Communication.
- Bell Syst. Tech. Jour., 27,
  juil. 1948 : pp 379–423
  oct. 1948 ; pp 623–656

[I - 1]
- **CG**
- Centre de Géostatistique
  35 rue St Honoré, 77305 Fontainebleau Cedex. France.
- Ecole Nationale Supérieure des Mines de Paris.

[I - 2]
- **Département de Géologie**
- 24 rue Lhomond
  75231 Paris, cedex 05, France.
- Ecole Normale Supérieure

[I - 3]
- **Département Géosciences Marines**
- Centres de Brest BP 70, 29263 Plouzané.
- Institut Francais de Recherche et d'Exploitation de la MER (IFREMER).

[I - 4]
- **CMM**
- Centre de Morphologie Mathématique
35 rue St Honoré, 77305 Fontainebleau Cedex. France.
- Ecole Nationale Supérieure des Mines de Paris.

Table des figures

Numéro	Titre	page
Chapitre 1		
1.1	Evolution dans le temps du taux de dioxyde de soufre sur la ville de BERLIN	18
1.2	Evolution dans le temps de la température de l'eau	19
1.3	Evolution dans le temps du magnétisme terrestre à l'observatoire de HUANCAYO	19
1.4	Cellules microscopiques (image "polypiers")	21
1.5	Les anomalies spatiales du magnétisme et de la gravimétrie. Un exemple de structures spatiales "quasi" périodiques	22
Chapitre 2		
2.1	Variations de la composante horizontale du champ magnétique en fonction de la latitude	28
2.2	Evolutions dans le temps du champ magnétique total à l'observatoire de KAKIOKA (Japon)	30 - 31
2.3	Spectre d'amplitude des perturbations KAKIOKA	35
2.4	Variogramme expérimental des perturbations KAKIOKA (25 jours)	37
2.5	Variogramme expérimental des perturbations KAKIOKA (7 jours)	37
2.6	Variogramme expérimental des perturbations KAKIOKA (3 jours)	38
2.7	Variogramme expérimental des perturbations KAKIOKA (1 jours)	38
2.8	Spectre d'énergie des perturbations KAKIOKA	41
2.9	Composante périodique du variogramme KAKIOKA	43
2.10	Composante non périodique du variogramme KAKIOKA	43
2.11	Ajustement de la composante non périodique du variogramme KAKIOKA (3 jours)	44
2.12	Ajustement de la composante non périodique du variogramme KAKIOKA (500 minutes)	44
2.13	Ajustement du variogramme total KAKIOKA (3 jours)	45
2.14	Ajustement du variogramme total KAKIOKA (1 jour)	45
2.15	Ajustement du variogramme total KAKIOKA (500 minutes)	45
2.16	Composante lunaire du variogramme KAKIOKA (25 jours)	47
2.17	Spectre d'amplitude des perturbations HUANCAYO (sur jours 1 à 3)	49
2.18	Spectre d'amplitude des perturbations HUANCAYO (sur jours 3 à 6)	49
2.19	Variogramme expérimental des perturbations HUANCAYO (3 jours)	52
2.20	Spectre d'énergie des perturbations HUANCAYO	53
2.21	Choix du domaine d'ajustement et détermination du palier (HUANCAYO)	54
2.22	Ajustement de la période principale (HUANCAYO)	55
2.23	Ajustement de la période secondaire (HUANCAYO)	55
2.24	Ajustement de la période tierce (HUANCAYO)	55
2.25	Modèle et réalité (ajustement visuel, HUANCAYO)	57
2.26	Modèle et réalité (ajustement par TFD, HUANCAYO)	58
2.27	Evolutions dans le temps du champ magnétique total au capteur immergé à MOHNS (cercle polaire)	60
2.28	Spectre d'amplitude des perturbations MOHNS	61
2.29	Variogramme expérimental des perturbations MOHNS (1.5 jours)	62
2.30	Spectre d'énergie des perturbations MOHNS	63
Chapitre 3		
3.1	Route d'échantillonnage en Géophysique Marine	68
3.2	Décomposition de la route en "pseudo-profils"	69
3.3	Méthode de recherche automatique du voisinage (BLUEPACK)	73
3.4	Géométrie du voisinage de type $\mathcal{V}_1^x$ (spatial "aveugle")	75
3.5	Géométrie du voisinage de type $\mathcal{V}_2^x$ (spatial anisotrope)	75
3.6	Géométrie du voisinage de type $\mathcal{V}_3^x$ (spatial isotrope "sautant")	76

Numéro	Titre	page
3.7	Géométrie du voisinage de type $\mathcal{V}_4^{x\ t}$ (spatio–temporel)	77
3.8	Genèse d'un voisinage spatio–temporel (résumé)	78
3.9	Localisation géographique de la campagne KRAKATAO	81
3.10	Route du bateau (KRAKATAO)	81
3.11	Zone test et profils utilisés (KRAKATAO)	82
3.12	Evolutions dans le temps du magnétisme spatio–temporel z(x,t) (KRAKATAO)	82
3.13	Magnétisme spatio–temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes (KRAKATAO)	83
3.14	Localisation géographique de la campagne SEAPERC	86
3.15	Zone test et profils utilisés (SEAPERC)	87
3.16	Evolutions dans le temps du magnétisme spatio–temporel z(x,t) comparativement aux perturbations diurnes à HUANCAYO	87
3.17	Cartographie des anomalies spatiales y(x)=z(x,t)–D(t) où D(t) est la fonction perturbations diurnes à HUANCAYO	88
3.18	Magnétisme spatio–temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes (SEAPERC)	88
3.19	Localisation géographique de la campagne MOHNS	91
3.20	Zone test et profils utilisés (zone dite "globale", MOHNS)	92
3.21	Zone test et profils utilisés (zone dite "locale", MOHNS)	92
3.22	Magnétisme spatio–temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes (MOHNS)	93
3.23	Evolutions dans le temps du magnétisme spatio–temporel z(x,t) (MOHNS)	93
Chapitre 4		
4.1	Magnétisme spatio–temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes (KRAKATAO)	101
4.2	Perturbations estimées (premier modèle, KRAKATAO)	102
4.3	Spectre de l'estimateur et post–filtrage (premier modèle, KRAKATAO)	103
4.4	Anomalies spatiales estimées (premier modèle, KRAKATAO)	104
4.5	Perturbations estimées (deuxième modèle, KRAKATAO)	105
4.6	Spectre de l'estimateur et post–filtrage (deuxième modèle, KRAKATAO)	106
4.7	Anomalies spatiales estimées (deuxième modèle, KRAKATAO)	107
4.8	Perturbations estimées (troisième modèle, KRAKATAO)	108
4.9	Spectre de l'estimateur et post–filtrage (troisième modèle, KRAKATAO)	109
4.10	Anomalies spatiales estimées (troisième modèle, KRAKATAO)	110
4.11	Le chemin parcouru grâce au filtrage (KRAKATAO)	112
4.12	Magnétisme spatio–temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes (SEAPERC)	113
4.13	Perturbations diurnes D(t) enregistrées à l'observatoire fixe de Huancayo (Pérou)	114
4.14	Cartographie des anomalies spatiales y(x)=z(x,t)–D(t) où D(t) est la fonction perturbations diurnes à HUANCAYO	114
4.15	Estimateur et réalité des perturbations (premier modèle, SEAPERC)	116
4.16	Anomalies spatiales estimées (premier modèle, SEAPERC)	117
4.17	Estimateur et réalité des perturbations (deuxième modèle, SEAPERC)	118
4.18	Estimateur et réalité des perturbations (troisième modèle, SEAPERC)	120
4.19	Anomalies spatiales estimées (troisième modèle, SEAPERC)	121
4.20	Estimateur et réalité des perturbations (quatrième modèle, SEAPERC)	122
4.21	Echantillonnage spatial correspondant aux points "explosifs" (quatrième modèle, SEAPERC)	123
4.22	Echantillonnage temporel correspondant aux points "explosifs" (quatrième modèle, SEAPERC)	124

Numéro	Titre	page
4.23	Estimateur et réalité des perturbations (cinquième modèle, SEAPERC)	128
4.24	Le chemin parcouru grâce au filtrage (SEAPERC)	130
4.25	Le chemin qu'il reste à parcourir ... (SEAPERC)	131
4.26	Magnétisme spatio-temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes (zone globale, MOHNS)	133
4.27	Magnétisme spatio-temporel z(x,t) cartographié dans le plan sans correction des perturbations diurnes (zone locale, MOHNS)	133
4.28	Evolutions dans le temps du champ magnétique total au capteur immergé à MOHNS (cercle polaire)	135
4.29	Spectre d'énergie des perturbations MOHNS	135
4.30	Perturbations estimées (premier modèle, MOHNS)	136
4.31	Spectre d'énergie de l'estimateur (premier modèle, MOHNS)	137
4.32	Variogramme de l'estimateur (premier modèle, MOHNS)	137
4.33	Variogrammes de la réalité et de l'estimateur (premier modèle, MOHNS)	138
4.34	Anomalies spatiales estimées (premier modèle, MOHNS)	138
4.35	Perturbations estimées (deuxième modèle, MOHNS)	139
4.36	Anomalies spatiales estimées (deuxième modèle, MOHNS)	140
4.37	Perturbations estimées (troisième modèle, MOHNS)	140
4.38	Estimateur et réalité des perturbations (troisième modèle, MOHNS)	141
4.39	Spectre d'énergie de l'estimateur (troisième modèle, MOHNS)	142
4.40	Variogramme sur 5 jours de l'estimateur (troisième modèle, MOHNS)	142
4.41	Variogrammes de la réalité et de l'estimateur (troisième modèle, MOHNS)	143
4.42	Anomalies spatiales estimées (troisième modèle, MOHNS)	143
4.43	Perturbations estimées (quatrième modèle, MOHNS)	144
4.44	Anomalies spatiales estimées (quatrième modèle, MOHNS)	144
4.45	Estimateur et réalité des perturbations (cinquième modèle, MOHNS)	145
4.46	Variogrammes des estimateurs (modèles 3 et 5) et de la réalité (MOHNS)	146
4.47	S'agit-il d'un filtrage ou d'un bruitage ?	147
Chapitre 5		
5.1	Image initiale (visage de GANDHI)	151
5.2	Variogrammes de l'image "GANDHI"	152
5.3	Ajustements des variogrammes de l'image "GANDHI"	153
5.4	Simulation d'une perturbation vidéophonique (période perturbatrice : 120 pixels)	154
5.5	Simulation d'une perturbation vidéophonique (période perturbatrice : 512 pixels)	155
5.6	Restauration obtenue par krigeage trigonométrique (période perturbatrice : 120 pixels)	156
5.7	Estimateur temporel et réalité (période perturbatrice : 120 pixels)	157
5.8	Restauration obtenue par krigeage trigonométrique (période perturbatrice : 512 pixels)	158
5.9	Estimateur temporel et réalité (période perturbatrice : 512 pixels)	158
5.10	Contours estimés par analyse krigeante trigonométrique (période ciblée : 8 pixels)	160
5.11	Contours estimés par analyse krigeante classique (périodes ciblées : 0 à 8 pixels)	161
5.12	Création d'un moiré par voisinage inadapté (périodes ciblées : 8 pixels, diamètre : 16 pixels)	162
Annexe B		
b.1	Inférence des caractéristiques de la FAI-k, résumé des étapes	188
b.2	Inférence du degré polynomial ("boucle-k")	190
b.3	Inférence de la covariance généralisée K(h) ("boucle-K")	192

	Numéro	Titre	page
	b.4	Validation du modèle ("boucle-K*")	193
Annexe D			
	d.1	Détection de fréquences temporelles (KRAKATAO, voisinage du type $\mathcal{V}_2^x$)	204
	d.2	Détection de fréquences temporelles (KRAKATAO, voisinage du type $\mathcal{V}_3^x$)	206
	d.3	Détection de fréquences temporelles (KRAKATAO, voisinage du type $\mathcal{V}_4^{x\,t}$)	207
	d.4	Inférence de la covariance généralisée K(h) (KRAKATAO, conditionnellement au premier ensemble de paramètres)	209
	d.5	Inférence de la covariance généralisée K(h) (KRAKATAO, conditionnellement au deuxième ensemble de paramètres)	210
	d.6	Inférence de la covariance généralisée K(h) (KRAKATAO, conditionnellement au troisième ensemble de paramètres)	211
Annexe E			
	e.1	Séquencement de la route du bateau en "pseudo-profils" lâches (campagne SEAPERC)	215
	e.2	Séquencement de la route du bateau en "pseudo-profils" courts (campagne SEAPERC)	215
	e.3	Configuration spatiale des échantillons sélectionnés (géométrie du type $\mathcal{V}_2^x$, campagne SEAPERC)	216
	e.4	Configuration temporelle des échantillons sélectionnés (géométrie du type $\mathcal{V}_2^x$, campagne SEAPERC)	217
	e.5	Configuration spatiale des échantillons sélectionnés (géométrie du type $\mathcal{V}_3^x$, campagne SEAPERC)	219
	e.6	Configuration temporelle des échantillons sélectionnés (géométrie du type $\mathcal{V}_3^x$, campagne SEAPERC)	220
	e.7	Configuration spatiale des échantillons sélectionnés (géométrie du type $\mathcal{V}_4^{x\,t}$, campagne SEAPERC)	222
	e.8	Configuration temporelle des échantillons sélectionnés (géométrie du type $\mathcal{V}_4^{x\,t}$, campagne SEAPERC)	223
Annexe F			
	f.1	Détection de fréquences temporelles (MOHNS, zone globale, voisinage du type $\mathcal{V}_2^x$)	227
	f.2	Détection de fréquences temporelles (MOHNS, zone locale, voisinage du type $\mathcal{V}_2^x$)	228
	f.3	Détection de fréquences temporelles (MOHNS, zone globale, voisinage du type $\mathcal{V}_3^x$)	229
	f.4	Détection de fréquences temporelles (MOHNS, zone locale, voisinage du type $\mathcal{V}_3^x$)	230
	f.5	Synthèse sur la détection de fréquences temporelles	232
	f.6	Inférence de la covariance généralisée K(h) (MOHNS, conditionnellement au premier ensemble de paramètres)	233
	f.7	Inférence de la covariance généralisée K(h) (MOHNS, conditionnellement au deuxième ensemble de paramètres)	234
	f.8	Inférence de la covariance généralisée K(h) (MOHNS, conditionnellement au troisième ensemble de paramètres)	235
	f.9	Inférence de la covariance généralisée K(h) (MOHNS, conditionnellement au quatrième ensemble de paramètres)	236
	f.10	Synthèse sur l'inférence et la validation de K(h)	237

TABLE DES MATIERES

REMERCIEMENTS 3
RÉSUMÉ DE LA THÈSE 5
ABSTRACT OF THE THESIS 7

INTRODUCTION 9
Le filtrage des perturbations diurnes en Géophysique Marine 11
Les Phénomènes à Tendance Périodique 11
La dualité spatial-spectral 11
Organisation du document 12

CHAPITRE 1 : LES PHENOMENES A TENDANCE PERIODIQUE (PTP) 13
Résumé et plan 13
Introduction 15

1.1 Description 15
1.1.1 En termes spatiaux 15
1.1.2 En termes spectraux 16
1.1.3 En termes de Théorie du Signal 17

1.2 Exemples 17
1.2.1 Les champs de séries temporelles 17
1.2.2 Autres exemples 21
1.2.3 Conclusions sur la section 1.2 23

CHAPITRE 2 : ANALYSE DE PTP TEMPORELS (perturbations diurnes en magnétisme) 25
Résumé et plan 25
Introduction 27

2.1 Préliminaires 27
2.1.1 Description géophysique des perturbations en magnétisme 29
2.1.2 Sur la TFD de covariances ou variogrammes 29

2.2 Série temporelle à grande échelle d'analyse : Kakioka 32
2.2.1 Présentation des données 32
2.2.2 Courbe expérimentale de D(t) 32
2.2.3 Périodogramme de D(.) 32
2.2.4 Variogramme expérimental de D(.) 37
2.2.5 Spectre d'énergie de D(.) 39
2.2.6 Anatomie d'un ajustement 41
2.2.7 Conclusions sur Kakioka 46

2.3 Série temporelle à petite échelle d'analyse : Huancayo 48
2.3.1 Présentation des données 48
2.3.2 Courbe expérimentale de D(t) 48
2.3.3 Périodogrammes de D(t) 48
2.3.4 Variogramme expérimental de D(t) 52
2.3.5 Spectre d'énergie de D(t) 52
2.3.6 Un premier ajustement : visuel 54
2.3.7 Un second ajustement : par TFD 58
2.3.8 Conclusions sur Huancayo 59

2.4 Série temporelle à faibles oscillations : MOHNS 60
2.4.1 Présentation des données 60

2.4.2 Courbe expérimentale de D(t) 60
2.4.3 Périodogramme de D(t) 61
2.4.4 Variogramme de D(t) 62
2.4.5 Spectre d'énergie de D(t) et conclusions 62

2.5 Conclusions du chapitre 2 63

CHAPITRE 3 : ANALYSE DE PTP SPATIO-TEMPORELS (magnétisme) 65
Résumé et plan 65
Introduction 67

3.1 Spécificités du problème 67
3.1.1 Le problème des perturbations diurnes en Géophysique marine . 67
3.1.2 Particularités de cet échantillonnage spatio-temporel 68
3.1.3 Hypothèses 69
3.1.4 Modifications des boucles d'inférence 70
3.1.5 Voisinages 72

3.2 Analyse et modélisation des données de la campagne Krakatao 80
3.2.1 Présentation des données 80
3.2.2 Les modèles inférés 83

3.3 Modélisation des données de la campagne Seaperc 85
3.3.1 Présentation des données 85
3.3.2 Les modèles inférés 89

3.4 Analyse et modélisation des données de la campagne Mohns 90
3.4.1 Présentation des données 90
3.4.2 Les modèles inférés 94

3.4 Conclusions du chapitre 3 95

CHAPITRE 4 : ESTIMATION DE PTP SPATIO-TEMPORELS (Magnétisme) 97
Résumé et plan 97
Introduction 99

4.1 Contexte des expérimentations 99
4.1.1 Les géométries du voisinage 99
4.1.2 Le système linéaire utilisé pour l'estimation 100

4.2 Filtrage des données de la campagne KRAKATAO 101
4.2.1 Présentation des données 101
4.2.2 Les modèles disponibles 101
4.2.3 Estimation avec le premier modèle 102
4.2.4 Estimation avec le deuxième modèle 105
4.2.5 Estimation avec le troisième modèle 107
4.2.6 Conclusions sur Krakatao 111

4.3 Filtrage des données de la campagne SEAPERC 113
4.3.1 Présentation des données 113
4.3.2 Les modèles disponibles 115

4.3.3 Estimation avec le premier modèle 115
4.3.4 Estimation avec le deuxième modèle 118
4.3.5 Estimation avec le troisième modèle 119
4.3.6 Estimation avec le quatrième modèle 121
4.3.7 Estimation avec le cinquième modèle 128
4.3.8 Conclusions sur Seaperc 129

4.4 Filtrage des données de la campagne MOHNS 132
4.4.1 Présentation des données 132
4.4.2 Les modèles disponibles 132
4.4.3 Estimation avec le premier modèle 136
4.4.4 Estimation avec le deuxième modèle 139
4.4.5 Estimation avec le troisième modèle 140
4.4.6 Estimation avec le quatrième modèle 144
4.4.7 Modification du pas d'échantillonnage 145
4.4.8 Conclusions sur Mohns 146

4.5 Conclusions du chapitre 4 148

CHAPITRE 5 : AUTRES APPLICATIONS (imagerie) 149
Résumé et plan 149

5.1 Filtrage d'une image à balayage 151
5.1.1 Présentation du problème 153
5.1.2 Liens avec le problème des perturbations en magnétisme 153
5.1.3 Système linéaire utilisé 156
5.1.4 Résultats du filtrage de la période T=120 pixel 156
5.1.5 Résultats du filtrage de la période T=2 x 256 pixel 157
5.1.6 Conclusions sur l'image à balayage 158

5.2 Estimation de hautes fréquences spatiales d'une image 159
5.2.1 Modèle disponible 159
5.2.2 Systèmes linéaires utilisés 159
5.2.3 Estimation par Analyse Krigeante Trigonométrique (AKT) 160
5.2.4 Estimation par Analyse Krigeante (AK) 161
5.2.5 Nécessité d'un diamètre adapté 161
5.2.6 Conclusions sur l'estimation des contours 162

CONCLUSIONS 163
Sur le filtrage des perturbations diurnes en magnétisme 165
Sur la généralisation à d'autres PTP 170
Sur la complémentarité spatial-spectral 171
Sur le krigeage de coefficients de Fourier 172
Sur le caractère esthétique et fascinant du PTP 173

ANNEXES 177

A Outil d'analyse spectrale : la Transformée de Fourier Discrète 179
A.1 Construction de la TFD 179
A.2 En résumé 180
A.3 Application aux fonctions périodiques 181
A.4 Application à l'interpolation 181
A.5 Généralisation à 2 dimensions 181

A-6 Notations 181

B Outils géostatistiques d'analyse structurale 183
B.1 La fonction variogramme 183
B.2 Inférence du variogramme 185
B.3 La fonction covariance généralisée K(h) 185
B.4 Inférence des caractéristiques de la FAI-k 187

C Krigeage (s) 195
C.1 Le principe "Krigeage" 195
C.2 Krigeage en FAI-k-{fi} 196
C.3 Le Krigeage de dérives 198
C.4 L'analyse krigeante 200

D Détails de la modélisation des données de la campagne Krakatao 203
D.1 Séquencement en "pseudo-profils" 203
D.2 Inférence du voisinage , de "k" et des fréquences temporelles ... 203
D.3 Inférence et validation de K(h) 208
D.4 Conclusions 212

E Détails de la modélisation des données de la campagne Seaperc 214
E.1 Modèle 1 : $\mathcal{V}$ du type $\mathcal{V}_2^x$, T=24h 214
E.2 Modèle 2 : $\mathcal{V}$ du type $\mathcal{V}_2^x$, T=24h et 12 h 218
E.3 Modèle 3 : $\mathcal{V}$ du type $\mathcal{V}_3^x$, T=24h et 12 h 218
E.4 Modèle 4 : $\mathcal{V}$ du type $\mathcal{V}_4^{x\,t}$, T=24h et 12 h 221
E.5 Modèle 5 : $\mathcal{V}$ du type $\mathcal{V}_4^{x\,t}$, FAI-$\emptyset$ forcée, T=24h, 12h 224
E.6 Conclusions 224

F Détails de la modélisation des données de la campagne Mohns 226
F.1 Séquencement en "pseudo-profils" 226
F.2 Inférence de $\mathcal{V}$, "k" et {fi} 226
F.3 Inférence et validation de K(h) 232
F.4 Conclusions 236

NOMENCLATURE DES NOTATIONS, ABREVIATIONS ET MOTS-CLEFS 239

BIBLIOGRAPHIES ET REFERENCES 243

TABLE DES FIGURES 247

TABLE DES MATIÈRES 251