

# Synthèse de processus bivariés non gaussiens à mémoires longues

Antoine SCHERRER, Patrice ABRY

Laboratoire de Physique, UMR 5672, CNRS, Ecole Normale Supérieure de Lyon,  
46, allée d'Italie, 69364 Lyon cedex 7, France  
prénom.nom@ens-lyon.fr

**Résumé** – Nous proposons une procédure pour la synthèse de processus bivariés dont les distributions marginales et les fonctions d'auto- et inter-covariances sont prescrites a priori. Nous détaillons ici le calcul pour des lois marginales gamma (de paramètres d'échelle multiples de demi-entier). Nous illustrons cette procédure sur un processus à longue mémoire bivarié dont les composantes possèdent la propriété dite de *connectivité fractale*. Nous indiquons la possibilité d'étendre le calcul à d'autres distributions et commentons la difficulté de synthétiser des processus dont les distributions marginales appartiennent à des familles de loi différentes.

**Abstract** – A procedure is proposed for the synthesis of bivariate stochastic processes whose marginal distributions and auto- and inter- covariance functions are a priori prescribed. Calculations corresponding to the case of marginal gamma distributions (with scale parameter restricted to multiple of semi integer values) are detailed. This procedure is illustrated at work for the synthesis of long range dependent processes whose bivariate components possess the so-called *fractal connectivity* property. The potential extension to other distributions is commented, together with the difficulties encountered when trying to obtain processes whose components have marginal distributions chosen amongst different families of laws.

## 1 Introduction

**Motivation.** La multiplication des technologies permettant de réaliser des capteurs, de les produire à bas-coût, de les mettre en réseaux, induit pour principale conséquence le fait que, pour la plupart des applications modernes (par exemple, le trafic Internet, les grands réseaux d'agent en interactions, les écoulements hydrodynamiques, les rythmes du corps humain, . . .), les données à analyser sont naturellement multivariées. De plus, elles présentent souvent des distributions marginales non gaussiennes (processus de comptage toujours positifs, par exemple), parfois à ailes lourdes, et mélangent, dans leur structure de corrélation, des effets de mémoires longue et courte. L'analyse de telles données nécessite l'élaboration d'outils de traitement spécifiquement adaptés. Cependant, ces propriétés particulières sont susceptibles de rendre délicate l'étude théorique de leur performances. Cette difficulté peut être circonvenue par le recours à des simulations numériques (type Monte Carlo), qui peuvent, par exemple, donner accès expérimentalement aux intervalles de confiance des procédures d'estimation ou à la puissance de tests d'hypothèses. De la même façon, la planification du développement ou l'exploitation d'un réseau de capteurs nécessitent des évaluations de performance ou de pertinence, qui peuvent ne pas être accessibles analytiquement. Là encore, la simulation numérique constitue une alternative

utile. Dans ces deux cas, il est essentiel de savoir produire numériquement des réalisations de processus multivariés dont certaines propriétés statistiques reproduisent celles observées ou attendues sur les données réelles à analyser, et sont donc prescrites a priori. Nous proposons ici une procédure de synthèse pour des processus bivariés dont les propriétés statistiques de premier (lois marginales non gaussiennes) et second (auto- et inter-covariances) sont choisies a priori.

**État de l'art.** La synthèse de processus gaussiens univariés de covariance prescrite a priori peut être réalisée efficacement par la méthode dite de *circulant embedded matrix* (cf. e.g., [12, 5] pour les versions originales et [3] pour un article de revue). Pour atteindre des distributions marginales non gaussiennes, diverses approches ont été proposées : outre la technique générale, mais approchée et itérative, dite *méthode de rejet* [8], l'inversion directe de la distribution cumulative peut être utilisée, pour les cas où cette inversion est analytiquement calculable. Cette inversion peut être explicite comme dans le théorème de Price [9], ou bien déclinée dans une technique de substitution de l'innovation (voir, par exemple, [7] et les références qui y sont citées). Nous avons exploité cette dernière approche pour la modélisation du trafic Internet [10]. Pour les processus gaussiens multivariés, un algorithme, proposé par Chambers [4] en permet une synthèse approchée. Il n'existe pas, à notre connaissance, de synthèse

multivariée non gaussienne.

**Contributions.** Pour combler cette lacune, nous combinons ici l'utilisation de l'algorithme de Chambers à une extension au cas bivarié de la technique de substitution de variable. Nous détaillons le calcul de cette procédure de synthèse dans le cas où les distributions marginales sont des lois gamma, dont les paramètres d'échelle sont multiples de demi-entiers. La section 2 décrit la méthode de synthèse en détail, cette méthode est illustrée dans la section 3 pour des processus possédant la propriété dite de *connectivité fractale* [2, 11], puis ses extensions sont discutées dans la section 4.

## 2 Procédure de synthèse

**Objectif.** Notre objectif est de décrire une procédure théorique permettant de produire effectivement en pratique des réalisations de processus aléatoires bivariés  $\{Y_1(k), Y_2(k), k = 1, \dots, n\}$ , possédant pour distributions marginales des lois gamma de paramètres différents (resp.  $(\alpha_1, \beta_1)$  et  $(\alpha_2, \beta_2)$ ), des fonctions d'autocovariance  $\gamma_{Y_1} = \sigma_{Y_1}^2 \rho_{Y_1}$  et  $\gamma_{Y_2} = \sigma_{Y_2}^2 \rho_{Y_2}$  (possiblement) différentes et une fonction d'intercovariance notée  $\gamma_{Y_{12}}$ . Les moyennes et variances s'écrivent respectivement  $\mu_{Y_i} = \alpha_i \beta_i$  et  $\sigma_{Y_i}^2 = \alpha_i \beta_i^2$ , pour  $i = 1, 2$ . Nous nous restreignons au cas où les  $\alpha_i$  sont entiers ou demi-entiers. Notons  $\alpha = \min(\alpha_1, \alpha_2)$ .

**Transformation non linéaire.** Il est bien connu qu'une variable gamma (de paramètre d'échelle  $\alpha$  demi entier) s'obtient comme somme de carré de variables gaussiennes, indépendantes, centrées, réduites :

$$i = 1, 2, \quad Y_i(n) = \frac{\beta_i}{2} \sum_{m=1}^{2\alpha_i} X_{i,m}^2(n), \quad (1)$$

où les  $\{X_{i,m}(n), i = 1, 2, m = 1, \dots, 2\alpha_i\}$  sont des processus gaussiens, centrés, réduits, dont les fonctions d'autocorrélation sont de la forme  $\forall m, \rho_{X_{i,m}} \equiv \rho_{X_i}$ . De plus, à  $i$  fixé, les processus  $\{X_{i,m}, m = 1, \dots, 2\alpha_i\}$  sont indépendants entre-eux. Enfin, les paires  $\{X_{1,m}, X_{2,m'}\}$ , sont choisies indépendantes dès que  $m' \neq m$   $\rho_{X_{1,m}, X_{2,m}} \equiv 0$  et avec une intercorrélacion identique sinon :  $\rho_{X_{1,m}, X_{2,m}} \equiv \rho_{X_{12}}, m = 1, \dots, 2\alpha$ .

Cette transformation non linéaire de variables gaussiennes vers une loi cible est évidemment spécifique de la distribution marginale choisie et doit être adaptée pour d'autres choix (voir [7] et [6]).

Il s'agit maintenant d'être capable d'exprimer  $\gamma_{X_1} \equiv \rho_{X_1}$ ,  $\gamma_{X_2} \equiv \rho_{X_2}$  et  $\gamma_{X_{12}}$  en fonction des fonctions  $\gamma_{Y_1}, \gamma_{Y_2}, \gamma_{Y_{12}}$  prescrites a priori.

**Autocovariances.** Pour établir le lien entre  $\rho_{X_i}$  et  $\rho_{Y_i}$ , on utilise la méthode de substitution de l'innovation (voir, par exemple, [7]). Pour cela, on écrit :

$$X_i(n+k) = \rho_{X_i}(k)X_i(n) + Z_i(n, k), \quad (2)$$

où l'innovation  $Z \equiv Z_i(n, k)$  est une variable aléatoire gaussienne indépendante de  $X_i(n), \forall n$ . Il est immédiat de

vérifier que  $Z$  est de moyenne nulle et de variance  $\sigma_Z^2 = 1 - \rho_{X_i}^2(k)$ . En injectant dans  $\sigma_{Y_i}^2 \rho_{Y_i}(k) = \mathbb{E}Y_i(k+n)Y_i(n) - \mathbb{E}Y_i(n+k)\mathbb{E}Y_i(n)$ , les transformation (1) et substitution (2), un calcul long mais non difficile (cf. [7, 10]) permet de montrer que :

$$\rho_{X_i} \equiv \gamma_{X_i} = \sqrt{\rho_{Y_i}} = \sqrt{\frac{\gamma_{Y_i}}{\alpha_i \beta_i^2}}. \quad (3)$$

Notons que cette relation indique de façon essentielle que la fonction d'autocorrélation d'un processus à loi marginales gamma (dont le facteur d'échelle est demi-entier) est nécessairement non négative.

**Intercovariance.** Pour obtenir la relation entre  $\gamma_{X_{12}}$  et  $\gamma_{Y_{12}}$ , nous proposons d'étendre la méthode de substitution de l'innovation à :

$$X_1(n+k) = \rho_{X_{12}}(k)X_2(n) + Z_{12}(n, k). \quad (4)$$

On écrit donc  $X_1(n+k)$  comme un terme prédit par  $\rho_{X_{12}}(k)X_2(n)$  et une innovation  $Z \equiv Z_{12}(n, k)$  gaussienne et indépendante des  $X_i(n), i = 1, 2$ . De façon évidente,  $Z$  est de moyenne nulle et variance  $\sigma_Z^2 = 1 - \rho_{X_{12}}^2(k)$ . On injecte alors les transformation (1) et substitution (4) dans  $\gamma_{Y_{12}}(k) = \mathbb{E}Y_1(n)Y_2(n+k) - \mathbb{E}Y_1(n+k)\mathbb{E}Y_2(n)$ , pour obtenir :

$$\begin{aligned} \gamma_{Y_{12}}(k) &= \mathbb{E}(Y_1(n)Y_2(n+k)) - \mathbb{E}(Y_1)\mathbb{E}(Y_2) \\ &= \frac{\beta_1\beta_2}{4} \left( \sum_{m=1}^{2\alpha_1} \sum_{m'=1}^{2\alpha_2} \mathbb{E}(X_{1,m}^2(n)X_{2,m'}^2(n+k)) \right) - \alpha_1\alpha_2\beta_1\beta_2 \\ &= \frac{\beta_1\beta_2}{4} \sum_{m=1}^{2\alpha} \mathbb{E}X_{1,m}^2(n)X_{2,m}^2(n+k) \\ &\quad + \frac{\beta_1\beta_2}{4} \sum_{m, m' \neq m=1}^{2\alpha_1, 2\alpha_2} \mathbb{E}X_{1,m}^2(n)\mathbb{E}X_{2,m'}^2(n+k) - \alpha_1\alpha_2\beta_1\beta_2 \\ &= \frac{\beta_1\beta_2}{4} \sum_{m=1}^{2\alpha} \mathbb{E}X_{1,m}^2(n)X_{2,m}^2(n+k) - \frac{\alpha\beta_1\beta_2}{2}. \end{aligned}$$

On calcule ensuite :

$$\begin{aligned} &\mathbb{E}X_{1,m}^2(n+k)X_{2,m}^2(n) \\ &= \mathbb{E}(\rho_{X_{12}}(k)X_2(n) + Z_{12}(n, k))^2 X_{2,m}^2(n) \\ &= \rho_{X_{12}}^2(k) \underbrace{\mathbb{E}X_2^4(n)}_{=3} + \underbrace{\mathbb{E}X_2^2(n)}_{=1} \underbrace{\mathbb{E}Z_{12}^2(n, k)}_{=1-\rho_{X_{12}}^2(k)} \\ &\quad + 2\rho_{X_{12}}(k) \underbrace{\mathbb{E}X_2^3(n)}_{=0} \underbrace{\mathbb{E}Z_{12}(n, k)}_{=0} \\ &= 1 + 2\rho_{X_{12}}(k). \end{aligned}$$

On combine ces deux résultats pour obtenir :

$$\gamma_{X_{12}} = \sqrt{\frac{\gamma_{Y_{12}}}{\alpha\beta_1\beta_2}}. \quad (5)$$

On remarque que cette procédure ne permettra d'accéder qu'à des processus gamma bivariés dont l'intercorrélacion

est partout non négative.

**Procédure de synthèse.** Lorsque le nombre d'échantillon  $n$  et les paramètres  $\alpha_i, \beta_i$  sont choisis, on sélectionne des covariances cibles  $\rho_{Y_i}$  et  $\gamma_{Y_{12}}$ , compatibles avec les équations (3) et (5) et permettant ainsi de calculer les  $\rho_{X_i}$  et  $\gamma_{X_{12}}$ . Pour  $m = 1, \dots, 2\alpha$ , on utilise l'algorithme de Chambers [4], non décrit ici faute de place, pour fabriquer  $2\alpha$  processus gaussiens bivariés  $\{X_{1,m}, X_{2,m}\}$ ,  $m = 1, \dots, 2\alpha$  et intercorrélés selon  $\gamma_{X_{12}}$ . Par simplicité, supposons que  $\alpha_1 \geq \alpha_2$ , les  $2(\alpha_1 - \alpha_2)$  séries manquantes  $\{X_{1,m}, m = 2\alpha_2 + 1, \dots, 2\alpha_1\}$  sont générées par *circulant embedded matrix*. Enfin, la relation (1) est utilisée pour produire les  $Y_i, i = 1, 2$ . Cette procédure de synthèse, en MATLAB est disponible sur requête auprès du second auteur.

### 3 Illustration et validation

**Connectivité fractale.** Pour illustrer la procédure de synthèse, nous allons utiliser le modèle de processus bivarié de marginales gammas (différentes) et ayant une structure corrélation suivant le modèle dit de *connectivité fractale* proposé dans [2] et testé dans [11]. La connectivité fractale repose sur deux ingrédients-clés : les auto-covariances mélangent (à la manière d'un processus farima) courtes et longues mémoires (d'où le terme *fractal*) ; les inter-covariances sont définies par une fonction de cohérence qui est constante non nulle dans la limite des basses fréquences (d'où le terme de *connectivité*). Ce modèle est bien défini par ses (auto et inter) spectres :

$$\Gamma_{Y_1}(\nu) = |1 - e^{-j\nu}|^{-2\theta_1} \Gamma_{Y_1}^*(\nu)$$

$$\Gamma_{Y_2}(\nu) = |1 - e^{-j\nu}|^{-2\theta_2} \Gamma_{Y_2}^*(\nu)$$

$$\Gamma_{Y_{12}}(f) = \Omega (1 - e^{-jf})^{-\theta_{1,2}} (1 - e^{jf})^{-\theta_{2,1}} \Gamma_{Y_{12}}^*(\nu)$$

Les paramètres  $\theta_*$  varient dans  $[0, 0.5]$ , et contrôlent la longue mémoire du processus, les fonctions  $\Gamma^*(\nu)$  sont non-négatives, symétriques, tendent vers 1 dans la limite des fréquences nulles. Elles contrôlent donc le comportement des hautes fréquences (courte mémoire). La connectivité fractale correspond au cas particulier pour lequel  $\theta_{1,2} + \theta_{2,1} = \theta_1 + \theta_2$ . Dans ce cas, la fonction de cohérence tend vers une constante non nulle dans la limite des basses fréquences :  $C(\nu) = \Omega, |\nu| \rightarrow 0$ .

**Validation.** Nous avons synthétisé 512 réalisations d'un processus bivarié de paramètres :  $\alpha_1 = 2, \beta_1 = 1, \alpha_2 = 3, \beta_2 = 2, \Omega = 0.7, \theta_1 = 0.1, \theta_2 = 0.3, \theta_{1,2} = \theta_{2,1} = 0.2$  avec  $\Gamma_Y^*(\nu) \equiv 1, \forall \nu$ .

La figure 1 (graphe en bas à droite) superpose les distributions visées (trait pointillé) et estimées (trait plein) : l'accord est excellent, même lorsque, comme c'est le cas ici, les lois marginales ont des paramètres différents. L'application systématique de tests d'hypothèses (de  $\chi^2$ ), non

rapportés ici, valident la pertinence de cette synthèse pour les distributions marginales.

Pour l'étude des auto- et inter-covariances, puisque l'objectif est de produire des processus satisfaisants au modèle de *connectivité fractale* (présence de longues mémoires liées sur les auto- et inter-covariances), nous avons recours à une analyse en ondelette des séries temporelles. Nous calculons les diagrammes log-échelles (log de la puissance des coefficients d'ondelettes à l'échelle  $2^j$  vs. log de l'échelle), ces outils, abondamment décrits par ailleurs (voir par exemple [1]), ne sont pas rappelés ici. Les diagrammes obtenus en moyenne sur les 512 réalisations (trait plein) sont comparés, sur la figure 1, aux diagrammes théoriques (trait pointillé). La ligne du haut correspond à chacune des composantes du signal. L'intercovariance des signaux est représentée en bas à gauche. L'accord entre fonctions cibles et obtenues illustre la pertinence et la puissance de la procédure.

### 4 Discussion et perspective

**Autres lois marginales appartenant à une même famille.** Cette procédure est en cours d'extension pour le cas où d'autres distributions marginales sont visées. Quand les deux distributions marginales appartiennent à la même famille (avec des paramètres différents), les calculs qui seront publiés dans [6] et permettent d'obtenir des marginales parmi les lois, exponentielle uni- ou bi-latérale, chi2, log-normal, pareto et uniforme.

**Lois marginales appartenant à des familles différentes.** L'utilisation de cette procédure pour produire des processus bivariés corrélés dont la distribution marginale de chaque composante est prise dans une famille différente est plus délicate. Examinons simplement ici le cas où  $\{Y_1, Y_2\}$  suivent respectivement une loi normale et  $Y_2$  et une loi Gamma.  $\{Y_1, Y_2\}$  s'obtiennent directement à partir de  $Y_1 = \sigma_1 X_{1,1} + \mu_1$  et  $Y_2 = \frac{\beta}{2} \sum_{m=1}^{2\alpha} X_{2,m}^2$ , avec  $\gamma_{X_1} \equiv \gamma_{Y_1}/\sigma_1^2$  et  $\gamma_{X_2} \equiv \sqrt{\gamma_{Y_2}/\sigma_2^2}$ . Pour obtenir l'intercovariance  $\gamma_{X_{12}}$ , on utilise la technique de substitution de l'innovation, avec  $X_{1,1}(n) = \gamma_{X_{12}}(k)X_{2,1}(n+k) + Z(n, k)$ , pour obtenir :

$$\begin{aligned} \gamma_{Y_{12}}(k) &= \frac{\beta}{2} \left[ \mathbb{E}((\sigma(\gamma_{X_{12}} X_{1,2}(n+k) \right. \\ &\quad \left. + Z(n, k)) + \mu) X_{2,1}^2(n+k)) + (2\alpha - 1)\mu \right] - \mu\alpha\beta \\ &= \frac{\beta}{2} \left[ \mathbb{E}(\sigma Z(n, k) X_{1,2}^2(n+k)) + \mathbb{E}(\sigma \gamma_{X_{12}} X_{2,1}^3(n+k)) \right. \\ &\quad \left. + \mathbb{E}(\mu X_{2,1}^2(n+k)) + (2\alpha - 1)\mu \right] - \mu\alpha\beta \end{aligned}$$

Or,  $\mathbb{E}(Z(n, k)) = 0$  et  $\mathbb{E}(X_{2,1}^3(n+k)) = 0$ , ce qui conduit à  $\gamma_{Y_{12}}(k) = 0$ . On retrouve ainsi le fait que des variables aléatoires suivant des distributions respectivement normale et gamma ne peuvent être corrélées.

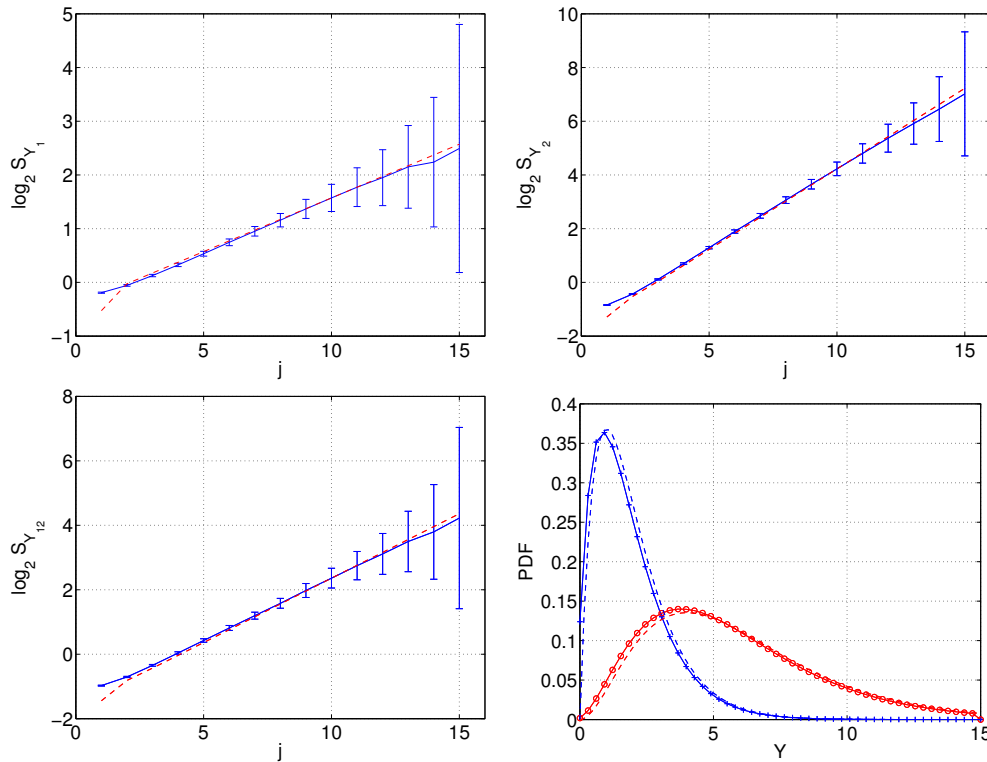


FIGURE 1 – **Illustration.** Comparaisons entre les diagrammes log-échelles théoriques (trait pointillé) et estimés à partir de moyennes sur des réalisations (trait plein) pour chacune des composantes du signal bivarié (en haut) et pour l'inter-covariance (en bas à gauche). Distributions marginales (en bas à droite) de chacune des composantes du signal.

Les calculs systématiques pour des distributions prises dans toutes les paires de famille citées ci-dessus seront publiés dans [6].

## Références

- [1] P. Abry, P. Flandrin, M. Taqqu, and D. Veitch, *Wavelets for the analysis, estimation, and synthesis of scaling data*, Self-similar Network Traffic and Performance Evaluation (New York) (K. Park and W. Willinger, eds.), Wiley, 2000, pp. 39–88.
- [2] S. Achard, D.S. Bassett, A. Meyer-Lindenberg, and E. Bullmore, *Fractal connectivity of long-memory networks*, *Phys. Rev. E* **77** (2008), no. 3, 036104.
- [3] J.-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, and M.S. Taqqu, *Generators of long range dependent processes : a survey*, Theory and applications of Long-range dependence (Boston) (P. Doukhan, G. Oppenheim, and M. S. Taqqu, eds.), Birkhäuser, 2003, pp. 579–624.
- [4] M.J. Chambers, *The simulation of random vector time series with given spectrum*, *Mathematical and Computer Modelling* **22** (1995), no. 6, 1–6.
- [5] C. R. Dietrich and G. N. Newsam, *Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix*, *SIAM Journal on Scientific Computing* **18** (1997), no. 4, 1088–1107.
- [6] H. Helgason, A. Scherrer, and P. Abry, *On a circulant embedded matrix synthesis procedure for non gaussian bivariate process*, preprint (2009).
- [7] S.B. Lowen and M.C. Teich, *Fractal-based point processes*, Wiley, Hoboken, NJ, 2005.
- [8] E. Merran, N. Hastings, and B. Peacock, *Statistical distributions*, John Wiley & sons, Inc., USA, 1993.
- [9] R. Price, *A useful theorem for non-linear devices having gaussian inputs.*, *IEEE Trans. Inform. Th.* **4** (1958), 69–72.
- [10] A. Scherrer and P. Abry, *Marginale non gaussienne et longue mémoire : analyse et synthèse de trafic internet*, Actes du 20ème Colloque GRETSI, 2005.
- [11] H. Wendt, A. Scherrer, P. Abry, and S. Achard, *Testing fractal connectivity in multivariate long memory processes*, Proceedings of ICASSP, 2009.
- [12] A. Wood and G. Chan, *Simulation of stationary gaussian processes in  $[0, 1]d$* , *Journal of computational and graphical statistics* **3(4)** (1994), 409–432.