Light cones and supervised learning prediction tasks

Imma Valentina Curato

Institute of Mathematical Finance, Ulm University

Ecodep Seminar, 02/05/2022

Satellites Data

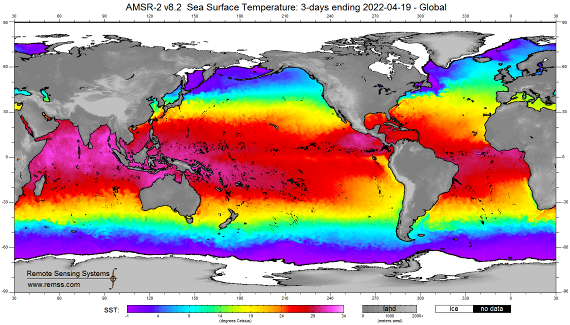


Figure: Global SST distribution derived by averaging satellites measurements over 3-days ending 19 April 2022, Available at www.remss.com/missions/amsr.

Satellites Data at different time stamps

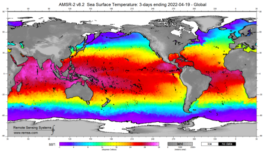


Figure: 3-Days ending 19/04/22

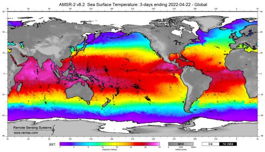


Figure: 3-Days ending 22/04/22

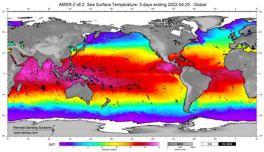


Figure: 3-Days ending 25/04/22

Satellites Data at different time stamps

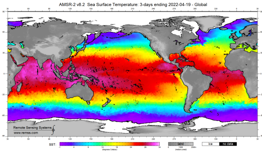


Figure: 3-Days ending 19/04/22

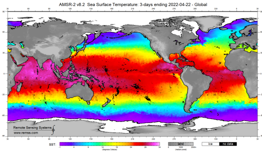


Figure: 3-Days ending 22/04/22

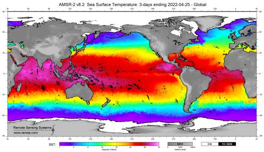


Figure: 3-Days ending 25/04/22

We can create a data set $\{Z_t(x)\}_{(t,x)\in\mathbb{T}\times\mathbb{L}}$, where $\mathbb{T} = \{t_i : i = 1, \dots, 3\}$ and $\mathbb{L} \subseteq \mathbb{R}^2$ is a regular spatial lattice representing the position of each pixel in the images (*frames*). Each value in the data set is representing the *color* of a pixel in a specific time stamp.

Light cones data

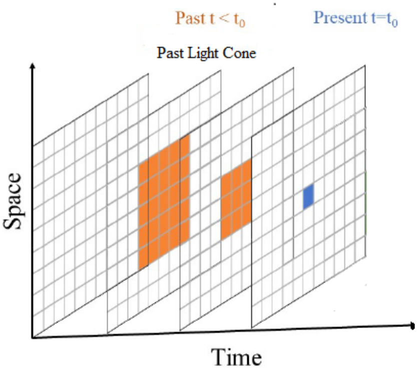
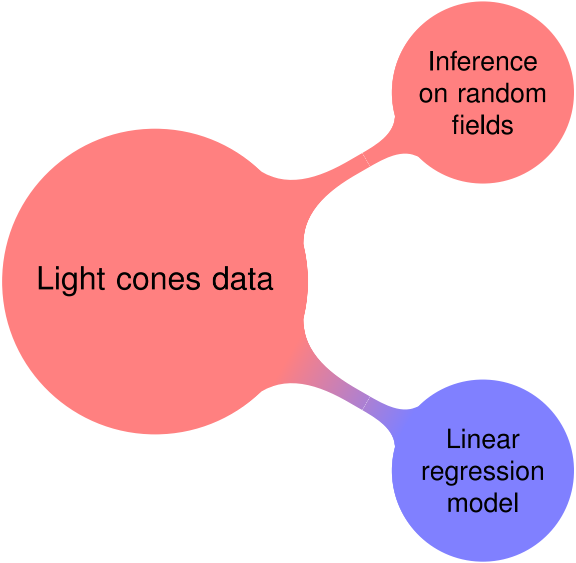


Figure: DiSco Project, Rupe et al. (2019).

$$L_2^-(t_0, x_0) = \{Z_{t_j}(x') : t_0 - 2 \leq t_j < t_0,\ x' \in \mathbb{L} \text{ such that } \|x' - x_0\| \leq c\,(t_0 - t_j)\}$$


Inference on random fields
Light cones data
Linear regression model

Moving average field defined on a cone

For a constant $c > 0$, let

$$A_t(x) = \{(s, \xi) \in \mathbb{R} \times \mathbb{R}^2 : s \leq t, \|x - \xi\| \leq c|t - s|\}.$$

Then, the random field

$$Z_t(x) = \int_{A_t(x)} f(x - \xi, t - s)\, \Lambda(d\xi, ds), \quad (t, x) \in \mathbb{R} \times \mathbb{R}^2,$$

where f is a deterministic function called *kernel* and Λ is a Lévy basis. $Z_t(x)$ is also called an **influenced moving average field** (MAF).

Influenced mixed moving average fields (MMAF)

$$Z_t(x) = \int_{\mathbb{R}} \int_{A_t(x)} f(A, x-\xi, t-s)\, \Lambda(dA, d\xi, ds), \quad (t,x) \in \mathbb{R} \times \mathbb{R}^2$$

Influenced mixed moving average fields (MMAF)

$$Z_t(x) = \int_{\mathbb{R}} \int_{A_t(x)} f(A, x - \xi, t - s)\, \Lambda(dA, d\xi, ds), \quad (t, x) \in \mathbb{R} \times \mathbb{R}^2$$

- A is a random parameter and its presence in the kernel function makes it possible to obtain short and long range dependence in space and time, see Nguyen and Veraart (2018).

Modeling light cones data

Each "color" observed in a pixel is influenced by its past light cone

$$\Updownarrow$$

$Z_t(x)$ is influenced by the values of the field in
$$L^-(t,x) = \{Z_{t'}(x') : (t',x') \in A_t(x)\}$$

Remark:

- The light cone describing the past of a specific pixel is a discretized version (analyzed for a finite time horizon and limited spacial grid) of the set $L^-(t,x)$.
- The light cone data set can be considered a sample from an influenced MMAF.

Spatio-temporal Ornstein Uhlenbeck field

For a constant $c > 0$, let

$$A_t(x) = \{(s, \xi) \in \mathbb{R} \times \mathbb{R}^2 : s \leq t, \|x - \xi\| \leq c|t - s|\},$$

$$Z_t(x) = \int_{A_t(x)} \exp(-\lambda(t - s))\, \Lambda(ds, d\xi), \quad (t, x) \in \mathbb{R} \times \mathbb{R}^2$$

Spatio-temporal Ornstein Uhlenbeck field

For a constant $c > 0$, let

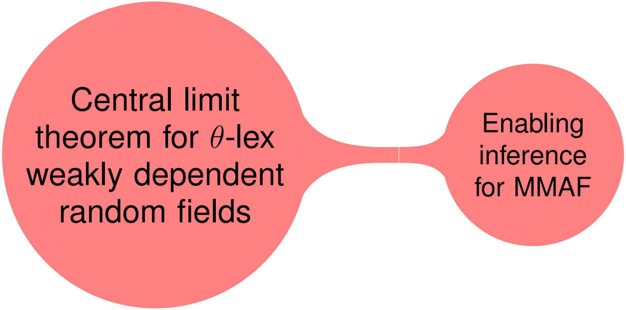
$$A_t(x) = \{(s, \xi) \in \mathbb{R} \times \mathbb{R}^2 : s \leq t, \|x - \xi\| \leq c|t - s|\},$$

$$Z_t(x) = \int_{A_t(x)} \exp(-\lambda(t - s)) \, \Lambda(ds, d\xi), \quad (t, x) \in \mathbb{R} \times \mathbb{R}^2$$

Can we define an estimator for the parameter λ?

GMM estimator

- C. and Stelzer (2019), prove asymptotic normality of the GMM estimator for a Lévy driven OU process and a Lévy driven supOU process.
- Nguyen and Veraart (2018), define a GMM for an STOU and an MSTOU field and study its consistency.


Central limit theorem for θ-lex weakly dependent random fields
Enabling inference for MMAF

Two ingredients

- *Stationary random fields*: a random field $(Z_t)_{t \in \mathbb{R}^d}$ is said to be strictly stationary if for any $n \in \mathbb{N}$, $\tau, t_1, \ldots, t_n \in \mathbb{R}^d$

$$(Z_{t_1+\tau}, Z_{t_2+\tau}, \ldots, Z_{t_n+\tau}) \stackrel{d}{=} (Z_{t_1}, \ldots, Z_{t_n}).$$

- *Strong mixing, Association, or Weak Dependence*[1] has to hold.

[1] See the monographs of Bradley (2007), Bulinskii and Shashkin (2007) and Dedecker et al. (2007) for a complete review.

A novel notion of dependence (C., Stelzer, and Ströh (2021))

θ-lex weak dependence

$\Downarrow$

A novel notion of dependence (C., Stelzer, and Ströh (2021))

θ-lex weak dependence

$\Downarrow$

If the field $\mathbb{E}[|Z_t|^\delta] \leq \infty$ for $\delta > 1$, then it is a more general notion of dependence than

- $\alpha_{\infty,1}$-mixing defined for random fields,

A novel notion of dependence (C., Stelzer, and Ströh (2021))

θ-lex weak dependence

$$\Downarrow$$

If the field $\mathbb{E}[|Z_t|^{\delta}] \leq \infty$ for $\delta > 1$, then it is a more general notion of dependence than

- $\alpha_{\infty,1}$-mixing defined for random fields,
- and α-mixing as defined for stochastic processes.

A novel notion of dependence (C., Stelzer, and Ströh (2021))

θ-lex weak dependence

$$\Downarrow$$

If the field $\mathbb{E}[|Z_t|^\delta] \leq \infty$ for $\delta > 1$, then it is a more general notion of dependence than

- $\alpha_{\infty,1}$-mixing defined for random fields,
- and α-mixing as defined for stochastic processes.

Influenced mixed moving average (MMAF) are θ-lex weakly dependent.

Lexicographic order

- For distinct elements $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ and $z = (z_1, \ldots, z_d) \in \mathbb{R}^d$ we say $y <_{lex} z$ if and only if $y_1 < z_1$ or $y_p < z_p$ for some $p \in \{2, \ldots, d\}$ and $y_q = z_q$ for $q = 1, \ldots, p-1$.
- Let $j \in \mathbb{R}^d$ and $h > 0$, we define

$$V_j^h = \{s \in \mathbb{R}^d : s <_{lex} j \text{ and } \|j - s\|_\infty \geq h\}.$$

Definition (C., Stelzer and Ströh)

A random field $Z = (Z_t)_{t \in \mathbb{R}^d}$ is θ-lex-weakly dependent if

$$\theta(h) = \sup_{u \in \mathbb{N}} \theta_u(h) \underset{h \to \infty}{\longrightarrow} 0,$$

where

$$\theta_u(h) = \sup \left\{ \frac{|Cov(F(Z_{i_1}, \ldots, Z_{i_u}), G(Z_j))|}{\|F\|_\infty Lip(G)}, j \in \mathbb{R}^d, \{i_1, \ldots, i_u\} \subset V_j^h, |\Gamma| = u \right\},$$

where F is a bounded function and G a bounded and Lipschitz function.

CLT for MMAF (C., Stelzer and Ströh)

Let $Z = (Z_t)_{t\in\mathbb{Z}^d}$ be a stationary centered real-valued MMAF such that $E[|Z_t|^{2+\delta}] < \infty$ for some $\delta > 0$. Additionally, assume that $\theta(h) \in \mathcal{O}(h^{-\alpha})$ with $\alpha > d(1 + \frac{1}{\delta})$. Then

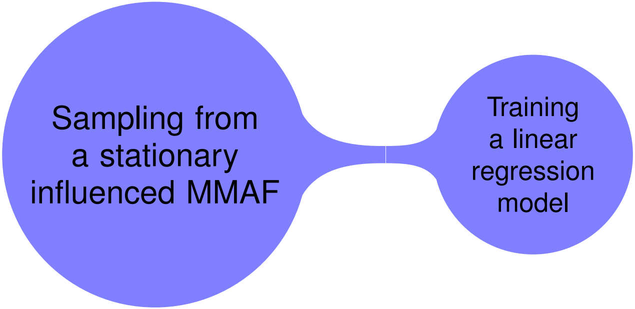
$$\sigma^2 = \sum_{k\in\mathbb{Z}^d} E[Z_0 Z_k],$$

is finite, non-negative and

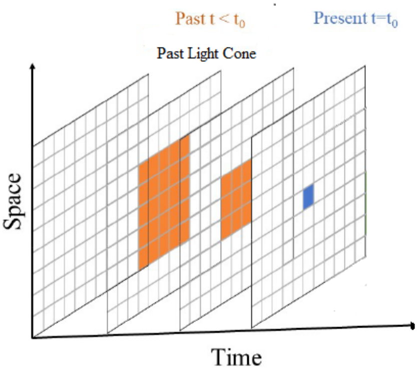
$$\frac{1}{|D_n|^{\frac{1}{2}}} \sum_{j\in D_n} Z_j \underset{n\to\infty}{\longrightarrow} N(0,\sigma^2),$$

where $(D_n)_{n\in\mathbb{N}}$ is a sequence of finite subsets of $\mathbb{Z}^d$ such that

$$\lim_{n\to\infty} |D_n| = \infty \text{ and } \lim_{n\to\infty} \frac{|\partial D_n|}{|D_n|} = 0.$$


Sampling from a stationary influenced MMAF
Training a linear regression model

Light cones data



Pre-processing data

- We define $S = \{(X_i, Y_i)\}_{i=1}^{m}$, where

$$X_i = L_p^-(t_i, x_i), \quad \text{and} \quad Y_i = Z_{t_i}(x_i).$$

$L_p^-(t_i, x_i)$ represents the *past light cone* of the point $(t_i, x_i) \in \mathbb{T} \times \mathbb{L}$ and is defined as the set of all points in the past (up to a finite horizon p) that could possibly influence the value $Z_{t_i}(x_i)$:

$$L_p^-(t_i, x_i) = \{Z_{t_j}(x') : t_i - p \leq t_j < t_i,\ x' \in \mathbb{L} \text{ such that } \|x' - x_i\| \leq c\,(t_i - t_j)\},$$

for $c > 0$.

Pre-processing data

- We define $S = \{(X_i, Y_i)\}_{i=1}^m$, where

$$X_i = L_p^-(t_i, x_i), \quad \text{and} \quad Y_i = Z_{t_i}(x_i).$$

  $L_p^-(t_i, x_i)$ represents the *past light cone* of the point $(t_i, x_i) \in \mathbb{T} \times \mathbb{L}$ and is defined as the set of all points in the past (up to a finite horizon p) that could possibly influence the value $Z_{t_i}(x_i)$:

$$L_p^-(t_i, x_i) = \{Z_{t_j}(x') : t_i - p \leq t_j < t_i,\ x' \in \mathbb{L} \text{ such that } \|x' - x_i\| \leq c\,(t_i - t_j)\},$$

  for $c > 0$.
- $\{(X_i, Y_i)\}_{i=1}^m$ are assumed to be realization of identically distributed random variables $(\boldsymbol{X}_i, \boldsymbol{Y}_i)_{i=1}^m$ which are jointly $\mathbb{P}$-distributed.

Aim: Linear regression task

Target: Determine a linear regression algorithm for light cones data, i.e., determining parameters $\beta = (\beta_0, \beta_1) \in B$ such that we determine a linear predictor

$$h_\beta(X) = \beta_0 + \langle \beta_1, X \rangle$$

.

Aim: Linear regression task

Target: Determine a linear regression algorithm for light cones data, i.e., determining parameters $\beta = (\beta_0, \beta_1) \in B$ such that we determine a linear predictor

$$h_\beta(X) = \beta_0 + \langle \beta_1, X \rangle$$

.

$$\Downarrow$$

An algorithm that predicts, e.g. the color, in pixel positions belonging to future images.

Supervised Learning

- The goal of a *learner* is to find a *predictor* h_β that minimizes the *generalization error* (also known as *theoretical risk*)

$$R(\beta) = \mathbb{E}[L(h_\beta(\boldsymbol{X}), \boldsymbol{Y})],$$

  where the *loss function* L is used to measure the discrepancy between a predicted output $h_\beta(\boldsymbol{X})$ and the real output $\boldsymbol{Y}$. This quantity is unobserved because we do not know the underlying distribution $\mathbb{P}$ of the data.

- We then select a predictor by minimizing *the empirical risk*

$$r(\beta) = \frac{1}{m}\sum_{i=1}^{m} L(h_\beta(X_i), Y_i).$$

- Generalization gap $= R(\beta) - r(\beta)$.

PAC ("Probably Approximately Correct") bounds for i.i.d. data

For all confidence parameters $\delta \in (0, 1)$, a generalization bound is defined such that

$$\mathbb{P}(\, \forall \, \beta \in B : R(\beta) \leq r(\beta) + \epsilon(\beta, m, \delta) \,) \geq 1 - \delta,$$

where $\epsilon(\beta, m, \delta)$ represents a measure of complexity and goes to zero as $m \to \infty$.

PAC ("Probably Approximately Correct") bounds for i.i.d. data

For all confidence parameters $\delta \in (0, 1)$, a generalization bound is defined such that

$$\mathbb{P}(\forall \beta \in B : R(\beta) \leq r(\beta) + \epsilon(\beta, m, \delta)) \geq 1 - \delta,$$

where $\epsilon(\beta, m, \delta)$ represents a measure of complexity and goes to zero as $m \to \infty$.

- May be computed just by using the **training data set**;
- Provide a computable control on the error on **any unseen data** with a specified confidence;
- **Algorithm:**
$$\hat{\beta} = arg \inf_{B} r(\beta) + \epsilon(\beta, m, \delta)$$

PAC (Generalized) Bayesian bounds for i.i.d data

Let $(B, \mathcal{T})$ a measurable space:

- Before seeing the data, fix a *reference* distribution π on B;
- Based on the data, learn a so called *generalized* distribution ρ on B;

PAC (Generalized) Bayesian bounds for i.i.d data

Let $(B, \mathcal{T})$ a measurable space:

- Before seeing the data, fix a *reference* distribution π on B;
- Based on the data, learn a so called *generalized* distribution ρ on B;

$$\mathbb{P}\Big(\forall\, \rho \,\mathrm{on}\, (B, \mathcal{T}) : \int R(\beta)\, d\rho(\beta) \leq \int r(\beta)\, d\rho(\beta) + \epsilon(\rho, \pi, m, \delta) \Big) \geq 1 - \delta.$$

Algorithm:

$$\hat{\rho} = arg \inf_{\rho} \int r(\beta)\, d\rho(\beta) + \epsilon(\rho, \pi, m, \delta)$$

PAC (Generalized) Bayesian bounds for i.i.d data

Let $(B, \mathcal{T})$ a measurable space:

- Before seeing the data, fix a *reference* distribution π on B;
- Based on the data, learn a so called *generalized* distribution ρ on B;

$$\mathbb{P}\left(\forall\, \rho \,\mathrm{on}\, (B, \mathcal{T}) : \int R(\beta)\, d\rho(\beta) \leq \int r(\beta)\, d\rho(\beta) + \epsilon(\rho, \pi, m, \delta) \right) \geq 1 - \delta.$$

Algorithm:

$$\hat{\rho} = arg \inf_{\rho} \int r(\beta)\, d\rho(\beta) + \epsilon(\rho, \pi, m, \delta)$$

Predictions: Draw $\beta \sim \hat{\rho}$ and predict with the chosen β.

PAC Bayesian bound for light cones data (C. and Ströh, 2022)[2]

- S is an identically distributed data set generated by a θ-lex weakly dependent random field with exponentially decaying coefficients $\theta_{lex}(h) \leq \bar{\alpha} \exp(-c^* h)$ for $c^* > 0$,
- $L^\epsilon = L \vee \epsilon$, for $0 < \epsilon < 3$,
- π be a distribution on B such that $\pi[\|\beta_1\|] \leq \infty$, then for all (regular conditional) distribution ρ such that $\rho << \pi$,
- $\alpha \in (0, 1)$,
- and KL represents the Kullback-Leibler divergence,

then for all $\delta \in (0, 1)$ and all ρ

$$\mathbb{P}\Big(\int R^\epsilon(\beta)\, d\rho(\beta) \leq \int r^\epsilon(\beta)\, d\rho(\beta) + \Big(KL(\rho||\pi) + \log\left(\frac{1}{\delta}\right) + \frac{3\epsilon^2}{2(3-\epsilon)}\Big) \frac{1}{m^{\frac{\alpha}{2}}} + \frac{1}{m^{\frac{\alpha}{2}}} \log\Big(\int 1 + 2(\|\beta_1\| + 1)\bar{\alpha}\, d\pi(\beta) \Big)\Big) \geq 1 - \delta$$

[2] This project is funded by the German Research Foundation, DFG GZ:CU 512/1-1

BRADLEY, R.C. (2007). Introduction to Strong Mixing Conditions, Volume 1, Kendrick Press, Utah.

BULINSKI, A. V. AND SHASHKIN, A. (2007). Limit theorems for associated random fields and related systems, World Scientific, Singapore.

CURATO, I. V., AND STELZER, R. (2019). Weak dependence and GMM estimation of supOU and mixed moving average processes. *Electron. J. Stat.* **13** 310–360.

CURATO, I., STELZER, R. AND STRÖH, B. (2021). Central limit theorems for stationary random fields under weak dependence with application to ambit and mixed moving average fields. *Forthcoming in Annals of Applied Probability.*

Curato, I. and Ströh, B. (2022). PAC Bayesian bounds for light cones data. *Working Paper.*

Dedecker, J. and Doukhan, P. and Lang, G. and León, J. R. and Louhichi, S. and Prieur, C. (2007). Weak dependence: with examples and applications, Springer-Verlag, New York.

Nguyen, M., and Veraart, A. E. D. (2018). Bridging between short-range and long-range dependence with mixed spatio-temporal Ornstein-Uhlenbeck processes. *Stochastics* **90** 1023–1052.

Rupe, A. and Kumar, N. and Epifanov, V. and Kashinath, K. and Pavlzk, O. and Schlimbach, F. and Patwary, M. and Maidanov, S. and Lee, V. and Prabhat and Crutchfield, J.P. (2019). DisCo: physics-based unsupervised discovery of coherent structures in spatiotemporal systems. *2019 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments* 75–87.

Thank you very much
for your attention

Ambit sets

$A_t(x) \subset \mathbb{R} \times \mathbb{R}^2$ such that

$$\begin{cases} A_t(x) = A_0(0) + (t, x), \text{ (Translation invariant)} \\ A_s(x) \subset A_t(x), \\ A_t(x) \cap (t, \infty) \times \mathbb{R}^2 = \emptyset. \text{ (Non-anticipative).} \end{cases}$$

Lévy Bases on $\mathbb{R}^q$ with mixing parameter in $\mathbb{R}$

The distribution of the Lévy basis is completely determined by the **characteristic quadruplet** $(\gamma, \Sigma, \nu, \pi)$. In particular,

$$E\left[e^{\mathrm{i}\langle u, \Lambda(B)\rangle}\right] = e^{\Phi(u)\Pi(B)}$$

for all $u \in \mathbb{R}^q$ and $B \in \mathcal{B}_b(\mathbb{R} \times \mathbb{R}^2)$, where $\Pi = \pi \times \lambda$ is the product measure of the probability measure π on $\mathbb{R}$ and the Lebesgue measure λ on $\mathbb{R}^2$. Furthermore,

$$\Phi(u) = \mathrm{i}\langle \gamma, u\rangle - \frac{1}{2}\langle u, \Sigma u\rangle + \int_{\mathbb{R}^q} \left(e^{\mathrm{i}\langle u,x\rangle} - 1 - \mathrm{i}\langle u, x\rangle 1_{[0,1]}(\|x\|)\right) \nu(dx)$$

is the cumulant transform of an ID distribution with characteristic triplet (γ, Σ, ν), where $\gamma \in \mathbb{R}^q$, $\Sigma \in M_{q\times q}(\mathbb{R})$ is a symmetric positive-semidefinite matrix and ν is a Lévy-measure on $\mathbb{R}^q$, i.e.

$$\nu(\{0\}) = 0 \quad \text{and} \int_{\mathbb{R}^q} \left(1 \wedge \|x\|^2\right) \nu(dx) < \infty.$$

Let $Z_t(x)$ be an MMAF driven by a Lévy basis with characteristic quadruplet $(\gamma, \Sigma, \nu, \pi)$ and with Λ-integrable kernel function $f : \mathbb{R} \times \mathbb{R}^2 \to M_{1\times q}(\mathbb{R})$.

- If $\int_{\|x\|>1} \|x\|\nu(dx) < \infty$ and $f \in L^1(\mathbb{R} \times \mathbb{R}^2, \pi \times \lambda) \cap L^2(\mathbb{R} \times \mathbb{R}^2, \pi \times \lambda)$ the first moment of X is given by

$$E[X_t] = \int_{\mathbb{R}} \int_{A_t(x)} f(A, t-s)\mu_\Lambda ds\pi(dA),$$

where $\mu_\Lambda = \gamma + \int_{\|x\|\geq 1} x\,\nu(dx)$.

- If $\int_{\mathbb{R}^q} \|x\|^2\nu(dx) < \infty$ and $f \in L^2(\mathbb{R} \times \mathbb{R}^2, \pi \times \lambda)$, then $X_t \in L^2$ and

$$Var(X_t) = \int_{\mathbb{R}} \int_{A_t(x)} f(A, t-s)\Sigma_\Lambda f(A, t-s)' ds\pi(dA) \text{ and}$$

$$Cov(X_0, X_t) = \int_{\mathbb{R}} \int_{A_0(x)\cap A_t(x)} f(A, -s)\Sigma_\Lambda f(A, t-s)' ds\pi(dA),$$

where $\Sigma_\Lambda = \Sigma + \int_{\mathbb{R}^q} xx'\nu(dx)$.

Let us consider an MMAF

$$X = (X_u)_{u \in \mathbb{Z}^d} \text{ with } X_u = \int_{\mathbb{R}} \int_{\mathbb{R}^d} f(A, u - s) \Lambda(dA, ds)$$

such that $E[X_0] = 0$ and $E[\|X_0\|^{2+\delta}] < \infty$ for some $\delta > 0$. Assume that X has η-coefficients satisfying $\eta_X(h) = \mathcal{O}(h^{-\beta})$, where $\beta > d \max\left(2, \left(1 + \frac{1}{\delta}\right)\right)$. Then,

$$\sigma^2 = \sum_{k \in \mathbb{Z}^d} E[Z_0 Z_k],$$

is finite, positive semidefinite and

$$\frac{1}{|D_n|^{\frac{1}{2}}} \sum_{u \in D_n} X_u \underset{n \to \infty}{\longrightarrow} N(0, \sigma^2).$$