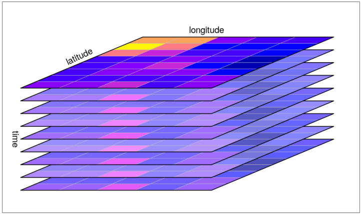
Mixed moving average field guided learning

Imma Valentina Curato
TU Chemnitz
in collaboration with
Orkun Furat, Lorenzo Proietti and Bennet Ströh

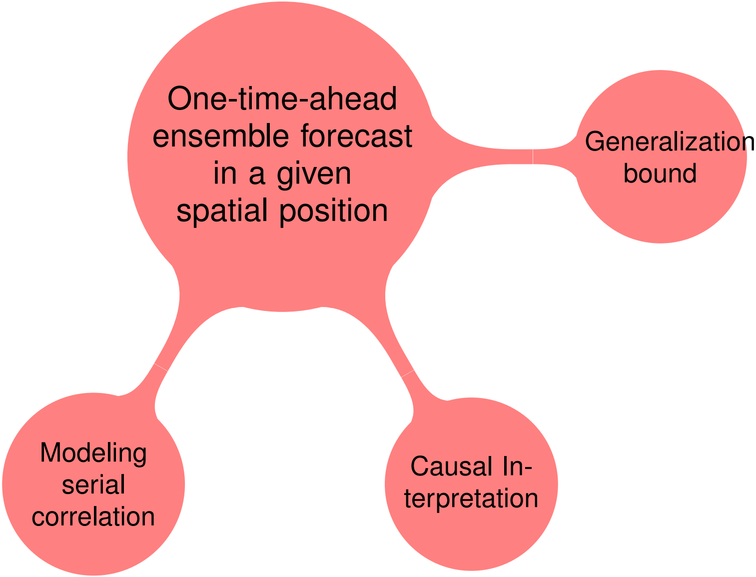
Multitask ECODEP Conference, IHP February 13th 2024

Raster Data Cube



- socio-economic or demographic data,
- environmental data
- time series of satellite images.

MMAF-guided learning



Theory-guided machine learning

1. We define a model underlying the data, i.e., a random field $\boldsymbol{Z} = (\boldsymbol{Z}_t(x))_{(t,x)\in\mathbb{R}\times\mathbb{R}^2}$ (for which we have no access to its predictive distribution, i.e. $\mathcal{L}_{\boldsymbol{Z}_{t_0}(x_0)|\boldsymbol{Z}_{t_1}(x_1),\ldots,\boldsymbol{Z}_{t_n}(x_n)}$);

Theory-guided machine learning

1. We define a model underlying the data, i.e., a random field $\boldsymbol{Z} = (\boldsymbol{Z}_t(x))_{(t,x)\in\mathbb{R}\times\mathbb{R}^2}$ (for which we have no access to its predictive distribution, i.e. $\mathcal{L}_{\boldsymbol{Z}_{t_0}(x_0)|\boldsymbol{Z}_{t_1}(x_1),\ldots,\boldsymbol{Z}_{t_n}(x_n)}$);
2. We employ properties of the underlying model to design a *generalized Bayesian algorithm*.

Causal Model for serially correlated spatio-temporal data

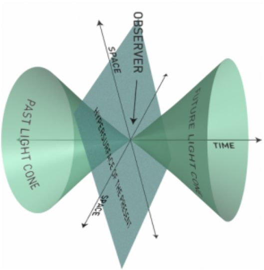


Figure: Past and future light cone

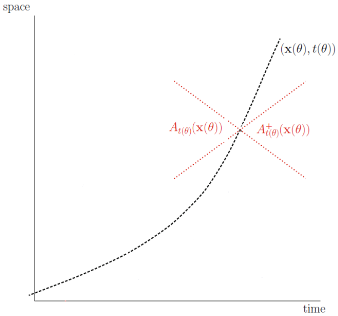


Figure: $A_t(x)$ is called an Ambit set (Barndorff-Nielsen et al. (2018)). Our methodology enables forecasts just in the space-time region $A_t(x)^+$.

Influenced Mixed 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

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

is called an influenced MMAF.

Influenced Mixed 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

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

is called an influenced MMAF.

- f is a deterministic function called *kernel* and Λ is a Lévy basis.
- 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).

Spatio-temporal Ornstein Uhlenbeck fields

Examples of MMAFs, are the STOU process

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

and its mixed version called MSTOU process

$$\boldsymbol{Z}_t(x) = \int_0^\infty \int_{A_t(x)} \exp(-\lambda(t-s))\,\Lambda(d\lambda, ds, d\xi), \quad (t,x) \in \mathbb{R}\times\mathbb{R}^2.$$

where λ is a random variable, typically described by a parametric distribution function.

Further Properties of MMAFs

Influenced Mixed moving average fields are:

- strictly *stationary*: i.e., for any $n \in \mathbb{N}$, $\tau, i_1, \ldots, i_n \in \mathbb{R} \times \mathbb{R}^2$,

$$(Z_{i_1+\tau}, Z_{i_2+\tau}, \ldots, Z_{i_n+\tau}) \stackrel{d}{=} (Z_{i_1}, \ldots, Z_{i_n});$$

- and θ-lex weakly dependent.

Asymptotic independence notions

- Strong Mixing, see Bradley (2007);
- Association, see Bulinskii and Shashkin (2007);
- Weak Dependence, see Dedecker et al. (2007).

Asymptotic independence notions

- Strong Mixing, see Bradley (2007);
- Association, see Bulinskii and Shashkin (2007);
- Weak Dependence, see Dedecker et al. (2007).
- **θ-lex weak dependence** is a novel dependence notion introduced in C., Stelzer and Ströh (2022) which extend to random fields the notion of θ-weak dependence introduced in Dedecker and Doukhan, "A new covariance inequality and applications", Stoch. Proc. Appl. (2003).

Lexicographic order

- For distinct elements $y = (y_1, y_2, y_3),\ z = (z_1, z_2, z_3) \in \mathbb{R} \times \mathbb{R}^2$ we say $y <_{lex} z$ if and only if $y_1 < z_1$ or $y_p < z_p$ for some $p \in \{2, 3\}$ and $y_q = z_q$ for $q = 1, \ldots, p - 1$.
- Let $j \in \mathbb{R} \times \mathbb{R}^2$ and $r > 0$, we define

$$V_j^r = \{s \in \mathbb{R} \times \mathbb{R}^2 : s <_{lex} j \text{ and } \|j - s\|_\infty \geq r\}.$$

θ-lex weak dependence

A random field $\boldsymbol{Z}_t(x)$ is θ-lex-weakly dependent if

$$\theta_{lex}(r) = \sup_{u \in \mathbb{N}} \theta_u(r) \underset{r \to \infty}{\longrightarrow} 0,$$

where

$$\theta_u(r) = \sup \left\{ \frac{|Cov(F(\boldsymbol{Z}_\Gamma), G(\boldsymbol{Z}_j))|}{\|F\|_\infty Lip(G)}, j \in \mathbb{R} \times \mathbb{R}^2, \Gamma \subset V_j^r, |\Gamma| = u \right\},$$

where F is a bounded function and G is a bounded and Lipschitz function. Moreover, $\Gamma = \{i_1, \dots, i_u\}$, and $\boldsymbol{Z}_\Gamma = (\boldsymbol{Z}_{i_1}, \dots, \boldsymbol{Z}_{i_u})$.

θ-lex weak dependence (C., Stelzer and Ströh (2022))

If the field ***Z*** admit finite moments $q > 1$, then it is a more general notion of dependence than

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

θ-lex weak dependence (C., Stelzer and Ströh (2022))

If the field $\boldsymbol{Z}$ admit finite moments $q > 1$, then it is a more general notion of dependence than

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

θ-lex weak dependence (C., Stelzer and Ströh (2022))

If the field $\boldsymbol{Z}$ admit finite moments $q > 1$, then it is a more general notion of dependence than

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

We will use the definition of ambit set and the θ-lex weak dependence of the underlying model to design our predictive algorithm.

Data decomposition

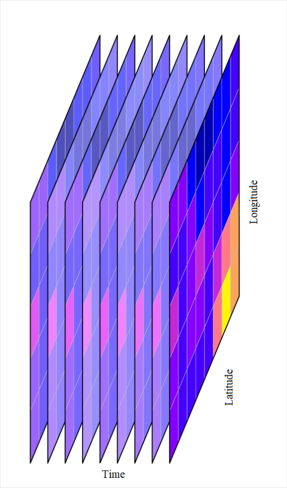


Figure: Raster Data Cube: observe data $\tilde{Z}_t(x) = \mu_t(x) + Z_t(x)$

Spatio-temporal embedding of N-frames

We aim to make *one-time ahead ensemble forecast* in a given spatial position x^* (supervised learning task), represented with a red pixel in the below picture.

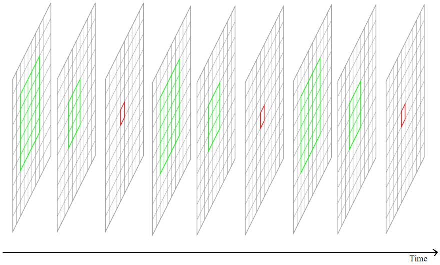


Figure: Exemplary training data set $S = \{(X_i, Y_i)^\top\}_{i=1}^m$ for $h_t = 1$, $c = \sqrt{2}, p_t = 2, a_t = 3, k = 1$. $X_i = L_p^-(t_0 + ia, x^*)$ (green pixels), with dimension $a(p, c) = 34$ and $Y_i = Z_t(x^*)$ (red pixel).

Discretized Ambit Set $\mathcal{I}(t, x^*)$

We define

$$X_i = L_p^-(t_0 + ia, x^*), \quad \text{and} \quad Y_i = Z_{t_0+ia}(x^*) \;\; \text{for } i = 1, \ldots, m,$$

where

- $L_p^-(t, x^*) = (Z_{i_1}(\xi_1), \ldots, Z_{i_{a(p,c)}}(\xi_{a(p,c)}))^\top$, and $(i_s, \xi_s) \in \mathcal{I}(t, x^*)$ for $s = 1, \ldots, a(p, c)$ and $t = t_0 + ia$ with $i = 1, \ldots, m$.
- We have that

$$\mathcal{I}(t, x^*) := \{(i_s, \xi_s) : \|x^* - \xi_s\| \leq c\,(t - i_s) \text{ for } 0 < t - i_s \leq p, \text{ and } (i_s, \xi_s) <_{lex} (i_{s+1}, \xi_{s+1})\},$$

  for $t = t_0 + ia$.

Discretized Ambit Set $\mathcal{I}(t, x^*)$

We define

$$X_i = L_p^-(t_0 + ia, x^*), \quad \text{and} \quad Y_i = Z_{t_0+ia}(x^*) \ \text{ for } i = 1, \ldots, m,$$

where

- $L_p^-(t, x^*) = (Z_{i_1}(\xi_1), \ldots, Z_{i_{a(p,c)}}(\xi_{a(p,c)}))^\top$, and $(i_s, \xi_s) \in \mathcal{I}(t, x^*)$ for $s = 1, \ldots, a(p, c)$ and $t = t_0 + ia$ with $i = 1, \ldots, m$.
- We have that

  $$\mathcal{I}(t, x^*) := \{(i_s, \xi_s) : \|x^* - \xi_s\| \leq c\,(t - i_s) \text{ for } 0 < t - i_s \leq p, \\ \text{and } (i_s, \xi_s) <_{lex} (i_{s+1}, \xi_{s+1})\},$$

  for $t = t_0 + ia$.
- The cone-geometry allows us to give a causal interpretation of the one-time ahead ensemble forecast.
- Moreover, $((\boldsymbol{X}_{i_1}, \boldsymbol{Y}_{i_1}), \ldots, (\boldsymbol{X}_{i_u}, \boldsymbol{Y}_{i_u}))$ and $(\boldsymbol{X}_j, \boldsymbol{Y}_j)$ for $u \in \mathbb{N}, i_1, \ldots i_u, j \in \mathbb{Z}$ and $i_1 \leq \ldots \leq i_u \leq j$ are lexicographically ordered marginals of the field $\boldsymbol{Z}$.

Theory-guided machine learning

1. We use an MMAF as the model underlying the data,

Theory-guided machine learning

1. We use an MMAF as the model underlying the data,
2. We employ the definition of ambit set and the asymptotic independence of the field to define a spatio-temporal embedding of the data,

Theory-guided machine learning

1. We use an MMAF as the model underlying the data,
2. We employ the definition of ambit set and the asymptotic independence of the field to define a spatio-temporal embedding of the data,
3. The latter is chosen in relation to the forecasting task we have in mind to perform.

Theory-guided machine learning

1. We use an MMAF as the model underlying the data,
2. We employ the definition of ambit set and the asymptotic independence of the field to define a spatio-temporal embedding of the data,
3. The latter is chosen in relation to the forecasting task we have in mind to perform.
4. S is a realization from the identically distributed random variables $\{(\boldsymbol{X}_i, \boldsymbol{Y}_i)^\top\}_{i=1}^m$ which are jointly $\mathbb{P}$-distributed and θ-weakly dependent.

Theory-guided machine learning

1. We use an MMAF as the model underlying the data,
2. We employ the definition of ambit set and the asymptotic independence of the field to define a spatio-temporal embedding of the data,
3. The latter is chosen in relation to the forecasting task we have in mind to perform.
4. S is a realization from the identically distributed random variables $\{(\boldsymbol{X}_i, \boldsymbol{Y}_i)^\top\}_{i=1}^m$ which are jointly $\mathbb{P}$-distributed and θ-weakly dependent.

Next, we use the training data set S in a supervised learning task.

Loss functions and hypothesis space

- $\mathcal{H}$ is the space of the Lipschitz functions: e.g., **linear functions**, **feed-forward neural networks**.
- Let $(\boldsymbol{X}, \boldsymbol{Y})^\top$ an input-output vector, and $L(h(\boldsymbol{X}), \boldsymbol{Y}) = |\boldsymbol{Y} - h(\boldsymbol{X})|$, the loss function used in the learning problem is

$$L^\epsilon(h(\boldsymbol{X}), \boldsymbol{Y})) = L(h(\boldsymbol{X}), \boldsymbol{Y})) \wedge \epsilon, \;\; \epsilon > 0.$$

- We define the **generalization error** as $R^\epsilon(h) = \mathbb{E}[L^\epsilon(h(\boldsymbol{X}), \boldsymbol{Y}))]$ and the **empirical error** as $r^\epsilon(h) = \dfrac{1}{m}\sum_{i=1}^{m} L^\epsilon(h(X_i), Y_i)$.

Parameters involved in MMAF-guided learning

Parameters	**Interpretation**	
h_t	*discretization step*	Observed
$a := a_t h_t$	*translation vector*	Chosen by the user
k	*further shift parameter*	Chosen by the user
$p := p_t h_t$	*past time horizon*	Hyperparameter
$m := \frac{N}{a}$	*number of examples in S*	Oberved+Derived
c	*speed of information propagation*	Estimated
λ	*decay rate of the θ-lex coef.*	Estimated
$a(p, c)$	*dimension of input-feature space*	Derived
ϵ	*accuracy level*	Hyperparameter

Generalized Bayesian setting

- Let π a probability measure on $\mathcal{H}$ that we call **generalized prior**.
- We aim to determine a conditional probability $\hat{\rho}$, called **generalized posterior** such that the average generalization gap

$$\int_{\mathcal{H}} R^{\epsilon}(h)\,\hat{\rho}(h) - \int_{\mathcal{H}} r^{\epsilon}(h)\,\hat{\rho}(h)$$

is small with high probability.

Fixed-time PAC Bayesian bound

Theorem

Let S be a training data sets generated by an MMAF field, $0 < \epsilon < 3$, $l = \left\lfloor \frac{m}{k} \right\rfloor$ *and* $r = ka - p$, π *be a distribution on* $\mathcal{H}$ *such that* $\pi[Lip(h)] \leq \infty$*. Then, for any* $\hat{\rho}$ *such that* $\hat{\rho} << \pi$*, and* $\delta \in (0, 1)$

$$\mathbb{P}\Bigg\{\left| \int_{\mathcal{H}} R^{\epsilon}(h)\,\hat{\rho}(h) - \int_{\mathcal{H}} r^{\epsilon}(h)\,\hat{\rho}(h) \right| \leq$$

$$+\left(KL(\hat{\rho}||\pi) + \log\left(\frac{1}{\delta}\right) + \frac{3\epsilon^2}{2(3-\epsilon)}\right)\frac{1}{\sqrt{l}}$$

$$+\frac{1}{\sqrt{l}}\log\left(\pi\left[1 + 2(Lip(h)a(p,c)+1)3\sqrt{l}\exp(3\sqrt{l})\theta_{lex}(r)\right]\right)\Bigg\} \geq 1-\delta.$$

Choosing the parameters a, k in the right way!

Let λ being the decay rate of the θ-lex coefficients of the field $\boldsymbol{Z}$, which we can estimate from the observed data,

$$\begin{cases} a_t > \dfrac{3\sqrt{N} + \log(3\sqrt{N}) + \lambda h_t p_t}{k h_t \lambda} & \text{if } \boldsymbol{Z} \text{ admits exponentially decaying } \theta\text{-lex coef.} \\ a_t > \dfrac{\exp(3\sqrt{N}/\lambda + \log(3\sqrt{N})/\lambda) + h_t p_t}{k h_t} & \text{if } \boldsymbol{Z} \text{ admits power decaying } \theta\text{-lex coef.} \end{cases}$$

then,

$$3\sqrt{l}\exp(3\sqrt{l})\theta_{lex}(r) \leq 1,$$

which gives us an idea on the order of magnitude of the addend in the PAC Bayesian bound.

Choosing the parameters a, k in the right way!

Let λ being the decay rate of the θ-lex coefficients of the field $\boldsymbol{Z}$, which we can estimate from the observed data,

$$\begin{cases} a_t > \dfrac{3\sqrt{N} + \log(3\sqrt{N}) + \lambda h_t p_t}{k h_t \lambda} & \text{if } \boldsymbol{Z} \text{ admits exponentially decaying } \theta\text{-lex coef.} \\[2ex] a_t > \dfrac{\exp(3\sqrt{N}/\lambda + \log(3\sqrt{N})/\lambda) + h_t p_t}{k h_t} & \text{if } \boldsymbol{Z} \text{ admits power decaying } \theta\text{-lex coef.} \end{cases}$$

then,

$$3\sqrt{l}\exp(3\sqrt{l})\theta_{lex}(r) \leq 1,$$

which gives us an idea on the order of magnitude of the addend in the PAC Bayesian bound.

The fastest convergence rate that can be obtained in this framework is $\mathcal{O}(m^{1/2})$ when choosing the parameter $k = 1$.

Any-time PAC Bayesian bound

Theorem

Let π be a distribution on $\mathcal{H}$ and S be a training data sets generated by an MMAF field. If $-\theta_{lex}^{Decay}(k) > 2\epsilon$ for $\epsilon > 0$, then for any $\hat{\rho} << \pi$, $m > 0$, and $\delta \in (0,1)$

$$\mathbb{P}\left\{ \left| \int_{\mathcal{H}} R^{\epsilon}(h)\,\hat{\rho}(h) - \int_{\mathcal{H}} r^{\epsilon}(h)\,\hat{\rho}(h) \right| \leq \left(KL(\hat{\rho}||\pi) + \log\left(\frac{1}{\delta}\right)\right)\frac{1}{\sqrt{m}} - \frac{1}{\sqrt{m}}\theta_{lex}^{Decay}(k) \right\} \geq 1 - 2\delta.$$

Details:

$$\theta_{lex}^{Decay}(k) := \begin{cases} \log(\exp(-\lambda h_t(ka_t - p_t)) & \text{if } \boldsymbol{Z} \text{ admits exponential decaying } \theta\text{-lex coef.} \\ \log((h_t(ka_t - p_t))^{-\lambda}) & \text{if } \boldsymbol{Z} \text{ admits power decaying } \theta\text{-lex coef.} \end{cases}$$

represents the decay of the exponential or power function appearing in the $\theta_{lex}(r)$ coefficient of the process $\boldsymbol{Z}$ for $r = ka - p$, where $a, p > 0$ and $k \in \mathbb{N}$, and

$$\begin{cases} a_t = \left\lceil \frac{2\epsilon}{k\lambda h_t} + \frac{p_t}{k} \right\rceil & \text{if } \boldsymbol{Z} \text{ admits exp. decaying } \theta\text{-lex coef.} \\ a_t = \left\lceil \frac{\exp(2\epsilon/\lambda)}{kh_t} + \frac{p_t}{k} \right\rceil & \text{if } \boldsymbol{Z} \text{ admits power decaying } \theta\text{-lex coef.} \end{cases}$$

Details:

$$\theta_{lex}^{Decay}(k) := \begin{cases} \log(\exp(-\lambda h_t(ka_t - p_t)) & \text{if } \boldsymbol{Z} \text{ admits exponential decaying } \theta\text{-lex coef.} \\ \log((h_t(ka_t - p_t))^{-\lambda}) & \text{if } \boldsymbol{Z} \text{ admits power decaying } \theta\text{-lex coef.} \end{cases}$$

represents the decay of the exponential or power function appearing in the $\theta_{lex}(r)$ coefficient of the process $\boldsymbol{Z}$ for $r = ka - p$, where $a, p > 0$ and $k \in \mathbb{N}$, and

$$\begin{cases} a_t = \left\lceil \frac{2\epsilon}{k\lambda h_t} + \frac{p_t}{k} \right\rceil & \text{if } \boldsymbol{Z} \text{ admits exp. decaying } \theta\text{-lex coef.} \\ a_t = \left\lceil \frac{\exp(2\epsilon/\lambda)}{kh_t} + \frac{p_t}{k} \right\rceil & \text{if } \boldsymbol{Z} \text{ admits power decaying } \theta\text{-lex coef.} \end{cases}$$

The convergence rate of this bound is $\mathcal{O}(m^{1/2})$ and we obtain the tightest version of the bound for $k = 1$.

Randomized Gibbs Estimator

Theorem (Oracle Anytime Bound)

Let π be a distribution on $\mathcal{H}$ such that $\bar{\rho} << \pi$ and

$$\frac{d\bar{\rho}}{d\pi} = \frac{\exp(-\sqrt{m}r^{\epsilon}(h))}{\pi[\exp(-\sqrt{m}r^{\epsilon}(h))]}.$$

If $-\theta_{lex}^{Decay}(1) > 2\epsilon$ for $\epsilon > 0$, then for any $\hat{\rho} << \pi$, $m > 0$, and $\delta \in (0, 1)$

$$\mathbb{P}\Big\{ \int_{\mathcal{H}} R^{\epsilon}(h)\,\bar{\rho}(h) \leq \inf_{\hat{\rho}} \Big(\int_{\mathcal{H}} R^{\epsilon}(h)\,\hat{\rho}(h) + \Big(KL(\hat{\rho}||\pi) + \log\Big(\frac{1}{\delta}\Big)\Big)\frac{2}{\sqrt{m}}\Big) - \frac{2}{\sqrt{m}}\theta_{lex}^{Decay}(1)\Big\} \geq 1 - 2\delta.$$

Data generated from a Spatio-Temporal Ornstein Uhlenbeck Process

Name	Mean Reverting Parameter	Lévy seed	Random generator seed
GAU1	$A = 1$	Gaussian	1
GAU10	$A = 4$	Gaussian	10
NIG1	$A = 1$	NIG	1
NIG10	$A = 4$	NIG	10

Table: Overview on simulated data sets with $N = 2000$, $c = 1$ and spatial dimension $d = 1$.

Inference on STOU process

Let $d = 1$ and $\boldsymbol{Z}$ be an STOU. If $\int_{|x|>1} x^2\, \nu(dx) \leq \infty$, $\gamma + \int_{|x|>1} x\, \nu(dx) = 0$, then $\boldsymbol{Z}$ is θ-lex weakly dependent with

$$\theta_{lex}(r) \leq \Big(\frac{c}{A^2} Var(\Lambda') \exp\Big(- \underbrace{\frac{A \min(2,c)}{c}}_{2\lambda} r\Big)\Big)^{\frac{1}{2}}$$

$$= \sqrt{2Cov(\boldsymbol{Z}_0(0), \boldsymbol{Z}_0(r \min(2,c)))} := \bar{\alpha} \exp(-\lambda r)$$

where $\lambda > 0$ and $\bar{\alpha} > 0$.

Inference on STOU process

Let $d = 1$ and $\boldsymbol{Z}$ be an STOU. If $\int_{|x|>1} x^2\, \nu(dx) \leq \infty$, $\gamma + \int_{|x|>1} x\, \nu(dx) = 0$, then $\boldsymbol{Z}$ is θ-lex weakly dependent with

$$\theta_{lex}(r) \leq \Big(\frac{c}{A^2} Var(\Lambda') \exp\Big(- \underbrace{\frac{A \min(2,c)}{c}}_{2\lambda} r \Big)\Big)^{\frac{1}{2}}$$
$$= \sqrt{2 Cov(\boldsymbol{Z}_0(0), \boldsymbol{Z}_0(r \min(2,c)))} := \bar{\alpha} \exp(-\lambda r)$$

where $\lambda > 0$ and $\bar{\alpha} > 0$.

We estimate the parameters A, c using the method of moments and the parameter λ using a plug-in estimator, which ultimately give us the right choice for the parameter a_t.

Causal Forecast

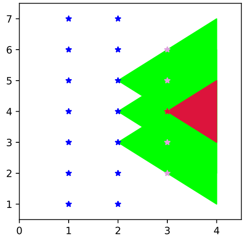
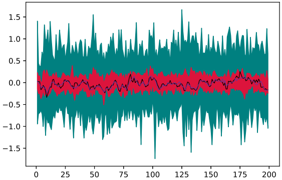
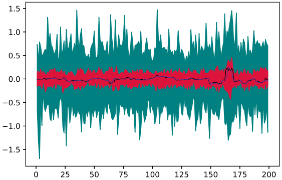


Figure: The x and y axes represent the time and spatial dimension, respectively. We picture the last two frames of a data set with spatial dimension $d = 1$ where the blue stars represent the pixels used in the definition of the training data set, and the violet stars represent the space-time points where it is possible to provide forecasts with MMAF guided learning for $p_t = c = h_t = 1$. Note that the forecast in the time-spatial position $(4, 3)$ lies in the intersection (red area) of the future lightcones $A_2(5)^+$, $A_2(4)^+$ and $A_2(3)^+$ and represented with

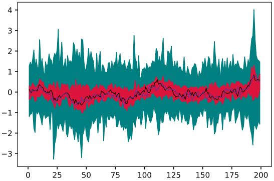
Ensemble Forecast using linear predictors: $p = 1, \epsilon = 3$



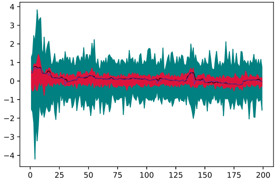
(a) GAU1A4



(b) NIG1A4



(c) GAU10



(d) NIG10

Thank you very much
for your attention

Bibliography

- BARNDORFF-NIELSEN, O. E. AND BENTH, F. E. AND VERAART, A. E. D. (2018). Ambit Stochastics, Springer, Cham.
- 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., STELZER, R. UND STRÖH, B. (2022). Central limit theorems for stationary random fields under weak dependence with application to ambit and mixed moving average fields. *Annals of Applied Probability* **32** 1814–1861.
- CURATO, I., FURAT, O. AND STRÖH, B. (2023). **Mixed moving average field guided learning for spatio-temporal data**, *arXiv:2301.00736.*

Bibliography

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.

DEDECKER, J., AND DOUKHAN, P. (2003). A new covariance inequality and applications *Stoch. Proc. Appl.* **106** 63-80.

NGUYEN, M., AND VERAART, A. E. D. (2017). Spatio-temporal Ornstein-Uhlenbeck processes:theory, simulation and statistical inference *Scand. J. Stat* **44** 46–80.

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.