Wasserstein GAN with data augmentation for the generation of multivariate extremes

Olivier Lopez, *Center for Research in Economics and Statistics, joint work with* ***Juliette Legrand***, ***Stéphane Lhaut***, ***Maud Thomas***

June 12th 2026

Outline

Outline

Wasserstein Generative Adversarial Networks

- GAN principle
- Generator
- The discriminator

Theoretical results

- Decomposition of the error
- Effect of the sample size
- Data augmentation

Illustration

- Simulation
- Real data example

Generative Adversarial Networks

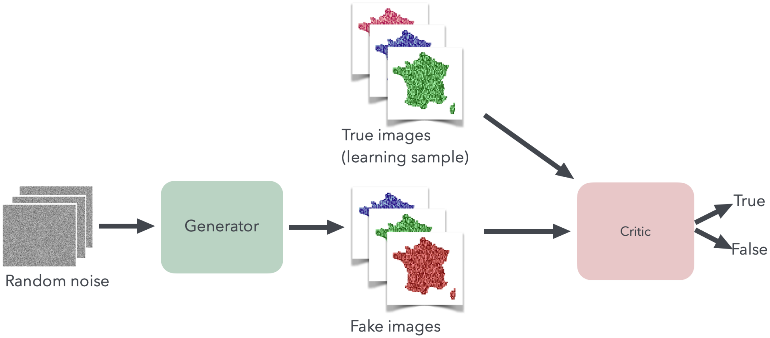


Figure 1 : Generative Adversarial Networks. The critic's objective is to detect fake images from real ones, while the Generator tries to « fool » the critic.

Multivariate GPD

- Multivariate GPD:
  - A multivariate GPD $\mathbf{Z} \in \mathbb{R}^d$ has the following representation,

$$\mathbf{Z} = \frac{\sigma}{\gamma}\left(\exp\left(\gamma\{E + \mathbf{S}\}\right) - 1\right),$$

where $E \sim \mathcal{E}(1)$ and $\mathbf{S} \in \mathbb{R}^d$ is independent from E with $\max(\mathbf{S}) = 0$.

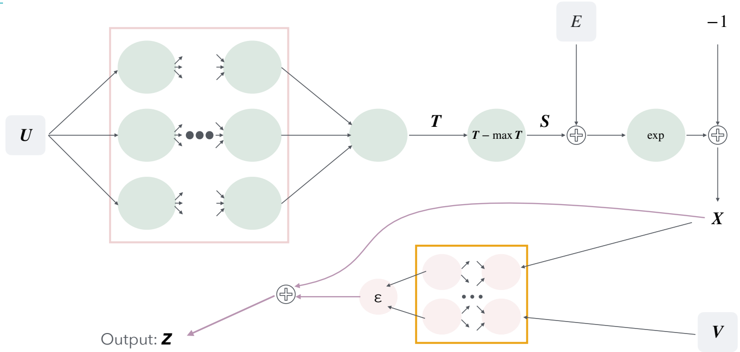
- For $\mathbf{Y}$ is the attraction domain of a distribution F_γ, there exists a function $\mathbf{s}(\mathbf{u})$ such that

$$\frac{(\mathbf{Y} - \mathbf{u}) \vee \mathbf{0}}{\mathbf{s}(\mathbf{u})} | \mathbf{Y} \not\leq \mathbf{u} \Longrightarrow \mathbf{Z} \vee \mathbf{0}, \ \mathbf{u} \to \infty.$$

- Observations:

$$(\mathbf{X}_i)_{1 \leq i \leq n} = \left(\frac{(\mathbf{Y}_i - \mathbf{u}) \vee \mathbf{0}}{\mathbf{s}(\mathbf{u})}\right)_{1 \leq i \leq n} | \mathbf{Y}_i \not\leq \mathbf{u}.$$

Generator



- Two parts: upper part produces a MGPD, the lower part captures misspecification.

Wasserstein distance

- Let us consider two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ in L^1.
- The Wasserstein distance between $\mathbb{P}$ and $\mathbb{Q}$ can be defined as

$$W(\mathbb{P},\mathbb{Q}) = \inf_{\pi \in \Pi(\mathbb{P},\mathbb{Q})} \int \|\mathbf{x}-\mathbf{y}\| \pi(d\mathbf{x}, d\mathbf{y}),$$

where $\Pi(\mathbb{P},\mathbb{Q})$ denotes the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with margins $\mathbb{P}$ and $\mathbb{Q}$.
- Dual version:

$$W(\mathbb{P},\mathbb{Q}) = \sup_{f \in Lip_1} \int f(\mathbf{x}) d\mathbb{P}(\mathbf{x}) - \int f(\mathbf{x}) d\mathbb{Q}(\mathbf{x}),$$

where Lip_1 denotes the 1$-$Lipschitz functions.

The discriminator and the Wasserstein distance

- Ideally, the discriminator (or critic) would compute $W(\mathbb{P}_\theta, \hat{\mathbb{P}})$, where $\hat{\mathbb{P}}$ is the empirical distribution of the examples, and $\mathbb{P}_\theta$ the distribution of the generator with parameters θ.
- **But** the supremum over Lip_1 is not computable.
- The critic identifies to a neural network $D_\alpha \in Lip_1$.
- It is designed so that

$$\sup_\alpha \int D_\alpha(\mathbf{x}) d\{\mathbb{P}_\theta - \hat{\mathbb{P}}\}(\mathbf{x}) \approx \sup_{f \in Lip_1} \int f(\mathbf{x}) d\{\mathbb{P}_\theta - \hat{\mathbb{P}}\}(\mathbf{x}).$$

Final criterion

- Wasserstein distance requires to have distribution in L^1, which is not necessarily the case here.
- We rely on a transformation of the variables.
- We try to solve the following optimization problem:

$$\inf_\theta \sup_\alpha \mathbb{E}[D_\alpha(\log(1+G_\theta(\mathbf{U})))] - \frac{1}{n}\sum_{i=1}^{n} D_\alpha(\log(1+\mathbf{X}_i))$$

- Remark: the logarithmic transformation is used for convenience but other transformations may be used.

Outline

Wasserstein Generative Adversarial Networks

- GAN principle
- Generator
- The discriminator

Theoretical results

- Decomposition of the error
- Effect of the sample size
- Data augmentation

Illustration

- Simulation
- Real data example

Theoretical results

- Let $\mathbb{P}_{\hat{\theta}}$ denote the distribution of the generator after training.
- Aim: bound

$$W_L(\mathbb{P}_{\hat{\theta}}, \mathbb{P}^*) = \sup_{f \in Lip_1} \int f(\log(1+\mathbf{x}))d\{\mathbb{P}_{\hat{\theta}} - \mathbb{P}^*\}(\mathbf{x})$$

  where $\mathbb{P}^*$ is the true distribution of $\mathbf{X}$.
- Decomposition of the error:
  - Approximation error: the ability of the generator to approximate the distribution of the excesses (comes from the MGPD approximation theorem);
  - Ability of the discriminator to approximate the supremum over 1-Lipschitz functions (see Biau et al. 2020, and Anil et al. 2019);
  - A term involving

$$W_{\mathcal{D}}(\mathbb{P}^*, \hat{\mathbb{P}}) = \sup_{\alpha} \mathbb{E}_{\mathbb{P}^*}[D_\alpha(\log(1+\mathbf{X}))] - \frac{1}{n}\sum_{i=1}^{n} D_\alpha(\log(1+\mathbf{X}_i)).$$

Effect of the sample size

Theorem

Let Q denote the total number of parameters in the discriminator, q the number of layers.

$$r_n(Q,q,d) = \frac{Q^{3/2}\left(qd^{1/2}\log n + q^2\right)}{n^{1/2}}.$$

There exists universal constants c_1, c_2, c_3, c_4, and ρ depending on the distribution of $\mathbf{X}$, such that, for all $x \geq c_1 r_n(Q,q,d)$,

$$\mathbb{P}\left(W_{\mathcal{D}}(\mathbb{P}^*, \hat{\mathbb{P}}) \geq x\right) \quad \leq \quad 2\exp\left(-\frac{c_2 x^2}{r_n(Q,q,d)^2}\right) + \exp(c_3 d^{1/2})\exp\left(-\frac{c_4 n\rho x}{Qqd^{1/2}}\right).$$

- In other words, $E[W_{\mathcal{D}}(\mathbb{P}^*, \hat{\mathbb{P}})] = O\left(r_n(Q,q,d)\right).$

Sketch of the proof

- Truncate the empirical sum by considering only the observations such that $\mathbf{X}_i \leq n^\beta$ for some β.
- Apply sub-gaussian equalities for this truncated part of the sum.
- Deal with the highest observations with Chernoff inequality.
- The constant ρ comes from the fact that we need

$$\mathbb{E}\left[\|(1+\mathbf{X})\|_\infty^\rho\right] < \infty,$$

for some $\rho > 0$ to control this last part.

Data augmentation via spectral bootstrap

- Question: can we add B synthetic observations to improve the rate?
- Idea: fit γ and $\boldsymbol{\sigma}$, and rely on the MGPD representation

$$\mathbf{Z} = \frac{\boldsymbol{\sigma}}{\gamma}\left(\exp\left(\gamma\{E + \mathbf{S}\}\right) - 1\right).$$

- "Spectral Bootstrap" (Madhar, Legrand, Thomas, 2025): get pseudo versions $(\hat{\mathbf{S}}_i)_{1\leq i\leq n}$ from the sample, and bootstrap them to obtain $(\mathbf{S}^*_b)_{1\leq b\leq B}$.
- Then compute

$$\mathbf{Z}^*_b = \frac{\hat{\boldsymbol{\sigma}}}{\hat{\gamma}}\left(\exp\left(\hat{\gamma}\{E_b + \mathbf{S}^*_b\}\right) - 1\right),$$

where E_b is unit exponential.

Effect of the spectral bootstrap

- Mixing these pseudo-observations with the sample changes the rate:
  - a "stochastic" term which is of rate $r_{n+B}(Q, q, d)$;
  - a bias term which is typically

$$O\left(\frac{B}{n+B} \times (\mathfrak{B}(\mathbf{u}) + r_n(Q, q, d))\right),$$

  where $\mathfrak{B}(\mathbf{u})$ is the bias caused by the approximation by a MGPD at threshold $\mathbf{u}$.
- Remark: typically, B should be less than n and $O(n)$. Asymptotically speaking, it does not change the rate, but it plays a role in modifying the constant.

Outline

Simulation

- Model A: MGPD distribution. Training sample: 2000 excesses.

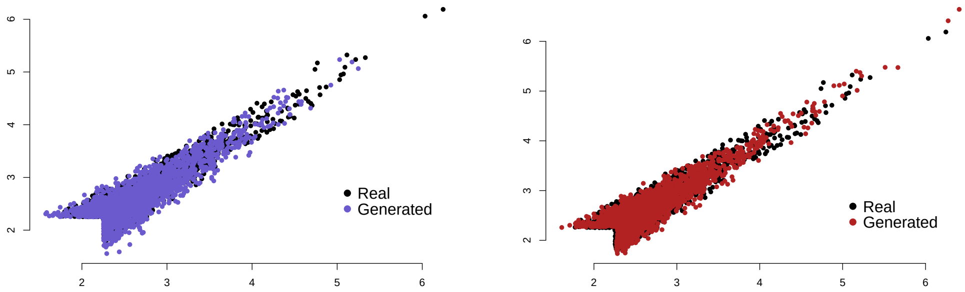


Figure: Model A. Left: MGP generator. Right: bias corrected.

Simulation

- Model B: Gumbel copula + Pareto margins. Training sample: 2000 excesses.

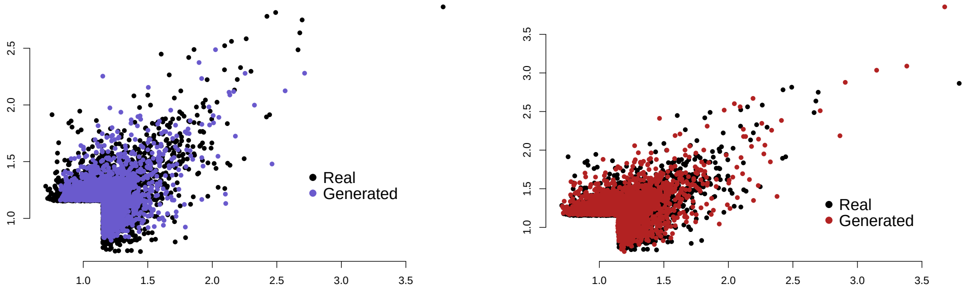


Figure: Model B. Left: MGP generator only. Right: bias corrected.

Financial data

- Portfolio of 10 industries over 26 years.
- On each margin, fit of an ARIMA + GARCH regime switching. We want to replicate the distribution of the residuals. $n = 1785$.

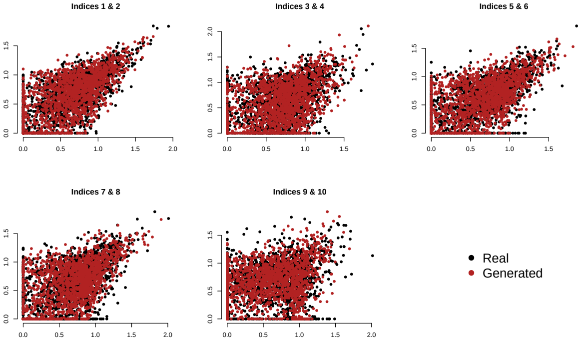
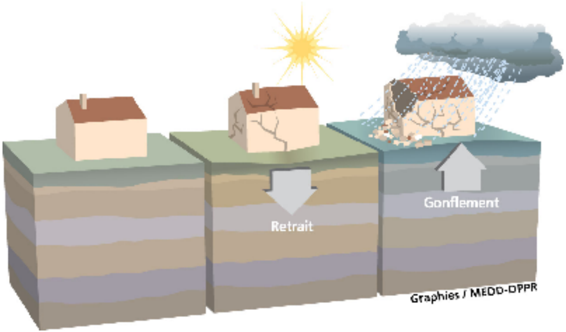


Figure: Generated bivariate adjacent pairs.

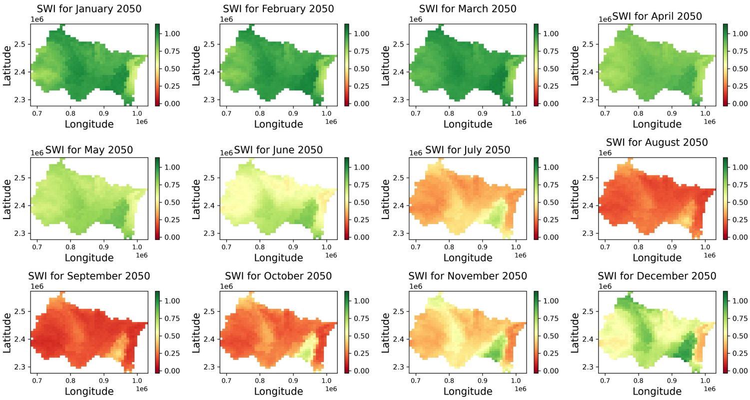
Conclusion

- Forthcoming: extension to insurance losses related to natural disasters.



Simulation of drought events

- In Ngatcha et al. (2026), frequency of drought episodes. Coming next: combination with severity via the access to insurance losses.



References

- Anil, C., Lucas, J., Grosse, R., *Sorting out Lipschitz function approximation*, International conference on machine learning (2021).
- Biau, G., Sangnier, M., Tanielian, U., *Some theoretical insights into Wasserstein GANs*, Journal of Machine Learning Research, 22(119), 1-45 (2021).
- Heranval, A., Lopez, O., Ngatcha, D., Nkameni, D. *A Wasserstein GAN-based climate scenario generator for risk management and insurance: the case of soil subsidence*, hal-05596862, 2026.
- Lhaut, S., Rootzén, H., Segers, J. *Simulation of multivariate extremes: A Wasserstein–Aitchison GAN approach*, Extremes (2026)
- Madhar, N., Legrand, J., Thomas, M., *Spectral Bootstrap for Non-Parametric Simulation of Multivariate Extreme Events*, Scandinavian Journal of Statistics (2026).

THANK YOU FOR YOUR ATTENTION

Center for Research in Economics and Statistics
CREST UMR 9194 - 5 avenue Henry Le Chatelier
91120 Palaiseau