PEPR EXPLORATOIRE TRACCS
APPEL À PROJETS
2024

PEPR TRACCS

DOCUMENT PRÉSENTATION PROJET

CHROMATIC

Acronyme du projet / Acronym	CHROMATIC
Titre du projet en anglais / Title in english	Machine learning for climate extreme impact
Titre du projet en français / Title in french	Apprentissage machine pour l'étude des impact des extrêmes climatiques
Mots clés / Keywords (min. 5 – max. 10)	distributional regression; extreme value theory; out-of-domain generalization; machine learning; coastal risks
Axes thématiques / Thematic axes	☐ Approches inter et transdisciplinaires pour le développement des services climatiques ☐ Evaluation des méthodes d'interventions climatiques X L'Intelligence Artificielle pour la modélisation du climat ☐ Evaluation des modèles de climat et/ou des services climatiques

Type d'instrument / Type of instrument	☐ Projet collaboratif / Collaborative project X Projet personnel de début de carrière / Early career fellowship

Montant envisagé de la subvention / Envisaged amount of the requested funding	350K €	
Durée envisagée du projet / Proposed project length	42 mois / months	
Responsable du projet / Coordinator of the project	**Nom, prénom, statut, établissement / Name, first name, status, organization**	
	BURITICA, Gloria, maître de conférences, AgroParisTech	
	Courriel / E-mail address	**Téléphone / Phone number**
	gloriapatricia.buriticaborda@agroparistech.fr	+33 07 51 05 33 61

	Dans le cas d'un projet personnel de début de carrière, veuillez indiquer : / For early career fellowships, please indicate:	
	Date de soutenance de thèse / PhD defense date	Date éventuelle de prise de fonction au sein d'un organisme ou établissement d'enseignement et de recherche / Start date of the current position in the organization, if applicable
	31/05/2022	01/09/2024
Etablissement coordinateur pressenti / Proposed leading institute	MIA Paris-Saclay, AgroParisTech	

English abstract

Extreme meteoceanic conditions can have devastating consequences on coastal areas, and for this reason it is important to identify the drivers of coastal risks and to quantify their impacts. A complete risk analysis of, for example, coastal flooding, requires assessing the effects and impacts of extreme climate conditions on this coastal hazard. Classical approaches to evaluate climate impact can impose strong assumptions between the drivers and the risk index variable. Instead, machine learning methods have the advantage of discovering complex relations in the data without assuming strong structural assumptions, and thus in many scenarios, they can reach outstanding performances in predicting the effect of drivers on a target variable. However, so far they are not tailored for predicting rare scenarios in the data including extreme events. Our proposal to the PEPR TRACCS: “Transforming climate modelisation for climate services" aims to investigate and develop advanced learning methods for correctly predicting the effects of extreme climate events to conduct reliable impact studies on the consequences of offshore forcing conditions like wave, sea-level, wind on coastal risks and human infrastructures.

One major difficulty with current machine learning algorithms is that they don't guarantee good predictions if a new predictor takes values that have rarely appeared in the training sample, or that have not yet been observed, and for this reason, they are not suitable for predicting extremes. Yet, in a climate change context, events that are rare today are likely to occur more frequently in the future, and for this reason, it is crucial to quantify the effects of extreme climate. This project plans to rely on extreme value theory to propose novel learning methods yielding good predictions under extreme scenarios. Regarding the methodology of our project,

1. First, we propose new algorithms for distributional regression relying on extrapolation techniques that provide satisfactory approximations of the effects of extreme predictors on the response variable (Rootzén, H. *et al.* 2018). Developing and evaluating their statistical guarantees will be the launching part of our project.

2. In a second part, these new developments will be applied to predict the extreme meteoceanic conditions in a data-driven way from climate variables (or any variables that are typical outputs of large-scale climate models, e.g., CFSR, ECMWF) that are of interest for impact assessment (e.g. extreme sea states at the coast) and impact indicators of interest for coastal risk management (e.g. maximum spatial flood extent, maximum water height at vulnerable assets).
3. Ultimately, the trained prediction models can be evaluated on different climate projections from numerical models (e.g. CMIP6 projections via Copernicus platform) to forecast the probable impacts of climate change.

To carry out this project we require a budget of 350K€ to hire a PhD student and an 18-month postdoc, and to cover the operational expenses of a regular collaboration between the members of the partnership. Moreover, we foresee strong interactions between our research direction and the PEPR TRACCS core project PC4-Extending, notably because the impact studies that we will conduct can help design risk plans in a climate change context. We also anticipate our methodology could help tackle further PC5-compact challenges regarding domain generalization. For example, for advancing hybrid AI climate models it is important to improve extreme event simulation when only a few examples of extreme events are available, and this topic also relates to the field of domain generalization.

French abstract

Les conditions météorologiques extrêmes peuvent avoir des conséquences dévastatrices sur les zones côtières. C'est pourquoi il est important d'identifier les facteurs de risques et de quantifier leurs impacts. Une analyse complète des risques, par exemple des inondations côtières, nécessite l'évaluation des effets des conditions climatiques extrêmes. Les approches classiques imposent des hypothèses fortes sur les covariables et la variable réponse. En revanche, les méthodes de machine learning sont capables d'apprendre des relations complexes sur les données sans supposer d'hypothèses structurelles fortes, et donc, dans de nombreux scénarios, elles peuvent atteindre des performances exceptionnelles de prédiction. Cependant, elles ne sont pas adaptées à la prédiction de scénarios rares, y compris les événements extrêmes. Notre proposition au PEPR TRACCS : « Transformer la modélisation du climat pour les services climatiques » vise à développer des méthodes d'apprentissage pour prédire correctement les effets des événements climatiques extrêmes afin de d'étudier les impacts des conditions de forçage offshore telles que les vagues, le niveau de la mer, le vent sur les risques côtiers et les infrastructures humaines.

L'une des principales difficultés des algorithmes de machine learning est qu'ils ne garantissent pas de bonnes prédictions si un nouveau prédicteur prend des valeurs qui sont rarement apparues dans l'échantillon d'apprentissage, et pour cette raison, ils ne sont pas adaptés à la prédiction des extrêmes. Pourtant, dans le contexte du changement climatique, les événements qui sont rares

aujourd'hui sont susceptibles de se produire plus fréquemment à l'avenir, et c'est pourquoi il est crucial de quantifier les effets du climat extrême. Ce projet prévoit de s'appuyer sur la théorie des valeurs extrêmes pour proposer de nouvelles méthodes d'apprentissage permettant d'obtenir de bonnes prédictions dans des scénarios extrêmes. En ce qui concerne la méthodologie de notre projet,

1. Premièrement, nous proposons de nouveaux algorithmes de régression en distribution reposant sur des techniques d'extrapolation en théorie des valeurs extrêmes (Rootzén, H. et al. 2018). Le développement et l'évaluation de leurs garanties statistiques constitueront la première partie de notre projet.
2. Dans une deuxième partie, ces nouveaux développements seront appliqués pour prédire les conditions météo-océaniques extrêmes à partir de données réelles sur des variables climatiques (ou variables des modèles climatiques à grande échelle, par exemple, CFSR, ECMWF) qui sont d'intérêt pour l'évaluation des impacts (par exemple, les états de mer extrêmes à la côte) et les indicateurs d'impact d'intérêt pour la gestion des risques côtiers (par exemple, l'étendue spatiale maximale des inondations, la hauteur d'eau maximale sur les biens vulnérables).
3. Enfin, les modèles de prévision proposés pourront être évalués sur différentes projections climatiques issues de modèles numériques (e.g., les projections CMIP6 via la plateforme Copernicus) afin de prévoir les impacts probables du changement climatique.

Pour ce projet, nous avons besoin d'un budget de 350K€ pour doctorant et un post-doctorant de 18 mois, et pour couvrir les dépenses opérationnelles d'une collaboration régulière entre les membres du consortium. Nous prévoyons de fortes interactions entre notre direction de recherche et le projet PC4-Extending du PEPR TRACCS, notamment parce que les études d'impact que nous mènerons peuvent aider à concevoir des plans de risque dans un contexte de changement climatique. Nous pensons également que notre méthodologie pourrait aider à relever d'autres défis du PC5-compact concernant la généralisation du domaine. Par exemple, pour faire progresser les modèles climatiques hybrides d'IA, il est important d'améliorer la simulation des événements extrêmes lorsque seuls quelques exemples d'événements extrêmes sont disponibles.

1 Context and challenges:

To anticipate the consequences of climate change, it is important to fully quantify impacts of extreme climate on society in terms of potential threats to people or infrastructures. The **CHROMATIC** project centers on developing machine learning (ML) prediction tools for conducting extreme impact studies. ML methods can learn complex relations between the predictors and the response if enough data is available. Yet, they can't extrapolate to scenarios with scarce data and are not reliable for prediction under extreme conditions. To enhance ML predictions in the extremes, we bring forward tools in extreme value theory. The applicability of field to tackle ML problems, particularly in regression, has only recently been studied and the few references focus either on extreme quantile regression (Gnecco, N. *et al.* 2023), or mean-squared loss minimization (Huet, N. *et al.* 2024) without considering distributional regression. Instead, building on the team's expertise in extreme value theory and on modeling extreme climate events' impacts on coastal assets, we propose to enhance distributional regression methods in ML to better anticipate the consequences of future extreme climate on coastal regions. Different application cases are envisioned, namely the survivability of on-offshore structures (offshore wind turbine, tidal turbine) and coastal risks (marine flooding, shoreline retreat) along the French Atlantic coast.

2 Project outline, methodology:

The **CHROMATIC** initiative aims to build reliable ML techniques to quantify extreme damage from extreme climatic conditions. The project will make use of data-based models (e.g. climate variables from CFSR, ECMWF among others) to train the proposed prediction methods, and impact studies will be carried out on simulated extremes from climate model projections (like CMIP6 from Copernicus platform). In terms of applications, we focus on predicting coastal risks and human infrastructure damages possibly caused by severe offshore meteoceanic conditions explained from large scale climate variables (e.g. sea level pressure, wind speed, sea surface temperature, etc). This project requires a PhD student in statistics and an 18-month postdoc in statistics/or climate sciences with strong programming skills to achieve the objectives that we detail next:

1. (Modeling the distributional effect of extreme events): High-impact extreme events can often be explained by other extreme predictors or drivers (e.g. extreme responses of offshore structures are driven by several meteoceanic drivers, Raillard, N. *et al*. 2018). This relation of extremal dependence has been largely studied in the field of extreme value theory. Relying on multivariate extreme value modeling (see e.g. Rootzén, H. *et al.* 2018), our first contribution will be to compute reliable approximations of the conditional distribution of a target variable based on predictors as these reach extreme records.

2. (Methodology development/evaluation): The first step typically entails restrictions on how extreme episodes occur and impact the response. The second step will be to propose regression methods in ML (adapting classical ones like random forests and neural networks), but tailored for learning extremal restrictions. Evaluating the method's capacity to learn these will be necessary to identify promising practices. Implementation and evaluation of the statistical guarantees of methods via numerical experiments and investigating consistency guarantees will be the second contribution of our work.

3. (Impact study of extreme climate scenarios): The first two steps will allow us to build data-driven prediction models. We then plan to conduct impact studies taking predictors from large-scale climate variables as CFSR, ECWM; and forecast target variables as extreme wave or sea coastal levels (data from hindcast simulations like HOMERE (IFREMER database)) or risk indicators like flooded area (numerically computed using hydrodynamics simulators). We focus on 1) meteoceanic conditions impacts at the coast (wave, sea level, wind); and 2) coastal risk impact on structures. The latter is more challenging as it implies translating the extreme coastal conditions into impact assessments indicators which is typically done using high-resolution numerical models. Combining the new developments with ML methods dedicated to this problem (see e.g. Rohmer et al., 2022) will be explored.

3 Description of the partnership

AgroParisTech (Gloria Buriticá) expert in extreme value statistics and ML tools for domain generalization, **IFREMER-French Marine Research Institute (Nicolas Raillard**) expert in statistical modeling of marine environments, **BRGM-French geological survey (Jérémy Rohmer)** expert in impact assessment for coastal applications using statistical tools and ML, **Université Paris Descartes (Anne Sabourin)** expert in extreme value theory, high dimensional statistics and statistical learning, project coordinator of the ANR-PRC 'EXSTA'.

1. Huet, N., Clémençon, Sabourin, A. (2024) On regression in Extremes. *arXiv:2303.03084*.
2. Gnecco, N., Terefe, E.D., Engelke S. (2023) Extremal random forests. *Journal of the American Statistical Association, Theory and Methods,* to appear.
3. Raillard N., Prevosto M., Pineau H. (2019) *3-D environmental extreme value models for the tension in a mooring line of a semi-submersible*. Ocean Engineering, 184, 23-31.
4. Rootzén H., Segers J., Wadsworth J.L. (2018) Multivariate generalized Pareto distributions: Parametrizations, representations and properties J. Multi. Anal. (165), 117-131.
5. Rohmer, J., Idier, D., Thieblemont, R., Le Cozannet, G., & Bachoc, F. (2022). Partitioning the contributions of dependent offshore forcing conditions in the probabilistic assessment of future coastal flooding. *Natural Hazards and Earth System Sciences*, 22(10), 3167-3182.

4 Positioning of the project in the general framework of the PEPR TRACCS

The **CHROMATIC** project develops the thematic axis 'AI for climate modeling' as we focus on developing novel practices in ML to enhance the prediction of the impacts and effects of extreme climate events. The hiring of a PhD student and postdoc will facilitate building and transferring knowledge aiming for a new generation of climate risk experts with acute statistical training and domain expertise. We anticipate interactions with the target project **PC4-Extending.** Indeed, the impact studies that we plan to conduct will help quantify the effect of climate change on target variables like total sea level at the coast or some flooding indicators, e.g. flood spatial extent, maximum water height at a given key asset. The conclusions of these studies can help develop contingency plans and build awareness of the consequences of climate change. We also foresee possible interactions with **PC5-Compact.** Indeed, the tools that we will develop could help solve further domain generalization problems in the field of AI. For example, they could be useful to improve generative algorithms aiming to simulate new extreme examples based on existing data. Advancing domain generalization tasks is key for transitioning from physical climate models to hybrid AI ones. Currently, most AI models are unreliable compared to physical climate models because they can't produce unprecedented but plausible extreme climate events. For this reason, addressing domain generalization problems in the field of AI is necessary to achieve this transition. Future collaborations with the PC5-Compact are possible in this line of research.