CY Initiative of Excellence

grant "Investissements d'Avenir" ANR- 16-IDEX-0008

funded

Project "EcoDep" CY-AAP2020-0000000013

The project *ecology_dependence* (EcoDep) aims at developing questions of ecology in which stochastic dependence is a main issue. A main ecological question is to qualify the development and the survival of the living species. The presence of a strong ecologist team in the staff will be essential to work out some main questions for the ecology of earth.

A strong collaboration of the team will be necessary to achieve our goals. Through our multi-disciplinary research team, Eco-Dep proposal aims to solve some of the current research challenges in the ecology in a world of complex environmental changes, global warming, and big data. Innovative aspects of the project are:

- Taylor's law for weakly-dependent time series and applications
- Population growth models and catastrophic events prediction and impact on the living populations
- Non-stationary time series and impact on the ecology and climate
- Causality, covariates and high dimension

For this a consortium of 11 researchers associated to CYU and to CY initiative, is essential ; they are mainly working with probabilistic and statistical tools. Among them, there are researchers from AGM, LPTM and THEMA (at Cergy-Pontoise University), Warwick and Mauritius (both associated to the CY Initiative).

A group of 4 ecologists outside of UCP will deeply contribute to the ecological thematics which are at the core of our project.

Now 7 more researchers in probability and statistics will contribute to the development of the necessary research items.

Three PhD students already working at UCP are also involved in the project.

Master level students will participate to a research dynamic

The project will need also 2 Postdocs with specific transverse skills, to be hired by Eco-Dep.

Our specific input will be to propose adapted dynamical models for ecological networks and to build coherent mathematical tools for their reconstruction.

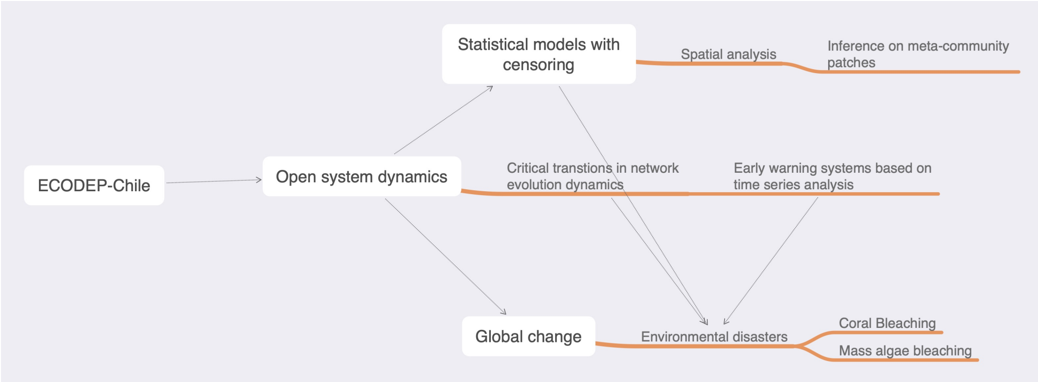
For this beyond a precise ecological approach one will need to adapt and improve on more classical statistical methods for stochastic processes.

Task 1 is dedicated to Tayor’s laws for population dynamics and their consequences.

Task 2 provides different stochastic modeling issues for ecological networks.

Task 3 makes it more precise how to derive inference for stochastic models for ecology.

After this statistical modeling task, one also needs to provide further characteristics of data sets. This is precisely the aims of the Task 4.



Severals task will be considered.

Task 1- **Taylor's dynamical laws**

Empirical studies show that the variance of the density of individual (per area or volume), on comparable populations, is a power-law function of the mean density of those populations.

This is referred to as the Taylor's law and it is confirmed in various species. This particular feature allows to understand the spatio-temporal variability of population dynamics, which is a central challenge for our problems. Indeed, as noted by Cohen et al. [CMPR12] «population fluctuations to low densities may increase risks of extinction, frustrating attempts at species conservation, and may cause genetic bottlenecks, which can have enduring evolutionary consequences. Fluctuations of fisheries, forests and agricultural crops have economic consequences and may directly affect human supplies of food, timber, fibre and fuel.» In this proposal, the Taylor's law will be considered in a dynamic of populations point of view, in collaboration with Cohen and together with Baydil, Doukhan, de la Pena, Prigent, Renault and Salhi.

The usual point of view concerning Taylor's laws of independent models has deep ecological applications as supported by [C1998]. Namely, a functional link relating the mean and the variance of non-negative random phenomena (see Cohen work).

Considering the standard limit of variance for sums (written as the series of covariances) instead of the individual variance allows to integrate the dependence structure and certainly exhibit new Taylor characteristics under dependence.

As noted before, Taylor's law means that the variance is proportional to its mean at some power α. Typically $\alpha=1$ or $\alpha=2$ are associated to Poisson or to Gaussian behaviours. This is at the basis of Cohen's work who already derived extremely successful conclusions for population dynamics. However, the dispersion is a characteristic of the marginal distribution of the involved phenomenon. Ongoing works in collaboration with Cohen, de la Pena, Baydil and Sahli, aim at replacing this marginal characteristic of dispersion by a time series analogous expression derived from the asymptotic properties of partial sums of weakly dependent time series.

In such cases, the variance of partial sums is equivalent to the product of the sample size by the sum of series of covariances. This sum appears as a reasonable measure for the dispersion of a time series.

This provides us with an alternative type of Taylor's law adapted to weakly dependent time series.

We expect more deductions for population dynamics in case such conditions are adapted.

For example, for many phenomena presenting an exponent $\alpha<2$ with respect to classical Taylor's law, it may be preferable to consider dynamic Taylor's law with exponent $\alpha=2$. It should help to understand the pattern and origins of fluctuations in population densities.

Better than considering the usual marginal characteristic based Taylor's law, this measure should be proved to be an efficient characteristic of the population dynamic evolution from a practical point of view. Extensive applications will be developed in this task and we will mainly consider the marine ecosystems.

Task 2- **Population growth models and Ecological networks**

Theoretical issues for modeling may be found in [CFM2006]. Population growth models subject to catastrophic events will be considered by Huillet, Goncalves, and Marquet. Especially some recent works by Marquet (see [B2012], [M2018]) plan terrible events in 2045 with ecological models. This would be a main topic to confirm through adequate probabilistic models.

We will study deterministic growth models punctuated by random catastrophes, either total (after such a catastrophe, the population size is set to 0 with probability 1...) or partial (after the catastrophe the population size randomly shrinks and is set to 0 with some probability in the interval [0,1)). The catastrophe rate is assumed state-dependent. Depending on state 0 being reflecting or absorbing, there is some interest for these semi-stochastic systems (as PDMPs) for example on: - the recurrence/transience transition criterion, - the shape of the invariant measure in the recurrent case, - the first return time to state 0, - the time to first extinction in general. Growth can be either polynomial, either sub-linear or super-linear, exponential or showing finite time explosion. The possibility of immigration is also considered. The project aims at giving a relevant formulation for different problems for population dynamics [BSZ2006] and[GH2019].

Ecological networks were modelled from a long time, see [M1972] for theoretical work on random interaction matrices, to empirical work on predator-prey (e.g. [M1992], [CBN2012]) and mutualistic networks (e.g. [B007]); their integration into *metanetworks* across space (e.g. [E2018]) and interaction type (trophic and non-trophic networks (e.g. [K015]) is also provided. This latter work was based on the ecological network of the Chilean intertidal ecosystem. Recently, the subtidal network considered in [P-M2017] will be useful in our theoretical studies of open networks and the impact of fisheries upon their resilience.

The inclusion of stochasticity, in ecological networks, within the open system paradigm, is a relatively recent development. In fact [R2019], [M2017] are among the pioneer works in this direction, showing that a Wright-Fisher diffusion dynamics in the flow of biomass among nodes (e.g. species, islands), appears as soon as one considers frequent perturbations of the main system.

Ecological and evolutionary views have focused on the mechanisms that enable or constrain species coexistence, genetic variation and the genetics of speciation, but a unified theory linking those approaches is still missing. There have been several ways to tackle this integration. For instance, Melian introduce evolutionary graphs in the context of neutral theories of molecular evolution and biodiversity to provide a framework that simultaneously addresses speciation rate and joint genetic and species diversities. This view coincides with current research in the open system approach in [R2019]. Finally a current research, carried by Videla and Rebolledo, focuses on the modeling of ecological networks submitted to variations on predators diets in a complex network.

In the project we are interested in the problem of **network reconstruction**, and estimation following a stochastic dynamics. That is, our main inference problem is connected with the reconstruction of a graph dynamics where the observables are proportions measured over the nodes of the graph. For instance, one may think that the graph represents populations of different species labelled by the nodes, so that the sum of respective proportions of biomass is one. As a result, the observables of our system are characterized as random variables taking values on the simplex. While the states, are the probability distributions of that random variables. Thus, we will first deal with probability distributions on the simplex. Numerous authors have been studying the special characteristics of these distributions as well as dependence of variables on the simplex (eg. [SR2002], [SR2002a]). This works aims at projection of species in the future.

Both continuous time and discrete time models are involved and their reconstruction is a main aim in the project. Models of interest may take values in unusual spaces as the above simplex. Note that continuous time models are usually discretely sampled so that we essentially defer this question to the next Task.

Task 3- **Non-stationary time series and impact on the ecology and climate**

Time series are e.g. given by population sizes measured annually, temperature data at some specified intervals, river-discharge, correlated spatial variables such as the variables measured along transects: plant cover, soil chemistry, water depth, see e.g. [Z2018] and [FNN2016]. More non-linear models are considered in [D2018].

The analysis of such dynamical data will moreover require the development of statistical models to measure the level of serial and cross-dependence and predict the frequency of occurrences. Mathematical measures of dependence are essential for their studies, as mixing [D1994] or weak dependence, see [DL1998] and [DDLLLP2007].

The global warming will be for instance considered through classes of models taking into account gradual [DRW2019], [T2019], [T2019a] and [BD2018], or structural changes [BKW2012] in the observations.

The assumption of monotonicity also provides data driven consistent procedures under such weak dependence conditions, as in [DLN2020], where a test for isotony is also introduced for integer valued models.

Other isotonic behaviours are natural also in the global warming scheme; adequate original methods associated with all those questions will thus be developed. Modeling of climatic extremes and natural disasters (hurricanes, typhoons), along with their spatial correlations and their effect on human populations.

Also questions about the possibility of insuring against such climatic extremes, are planned with Doukhan, Heinen, Prigent and Sahli.

Isotonic models are considered and estimated in [FLN2018], [DLN2020]. The assumption of monotoni-city indeed yields a self-tuned estimator with non-parametric minimax rate under additional mixing assumptions. This feature will be exploited in order to fit the precise intensity of the global warming.

We will then focus on the study of the dependence structure and statistical properties in order to detect and predict gradual changes in such model. Many possibilities can be considered from change point analyses to local stationarities.

The case of integer valued time series is of a special importance for the model of dynamic of populations [D2018] and will be developed with various mamber in Eco-Dep, namely, Doukhan, Heinen, Garnier, Mamode Khan, Neumann, and Sahli.

[BCD2019a] and [BCD2019b] propose the estimation of the derivative of the regression function in fixed and random designs stochastic dynamic regression. [DGPT2015] and [DPFM2016] also developed a new theoretical and applied framework for the extreme quantile estimation with application in ecology. In the present context of globally increasing sea surface temperature associated to climate change, we were able to detect in particular physiological impact of global warming in the IORO reef in the Havannah channel in the southern lagoon of New Caledonia. With tropical reefs around the world threatened by warming oceans, the effect of environmental conditions on giant clams was shown and we suggest that the giant clam Hippopus hippopus could be an interesting sentinel species. The combination of statistical dynamic modeling procedure including with high-frequency data measurement techniques will provide a good way for studying complex ecosystem.

Other related questions as the development of coral reefs and the prediction of marine ecosystems in the Chilean Pacific will include Bahamonde, Doukhan, Durrieu, Heinen, Kengne, Mamode Khan, Neumann, Truquet, Salhi, and Navarrete.

Task 4- **Causality, covariates and high dimension**

Covariates are important to consider for many studies, and they include non-stationary models especially adapted to deal with population dynamic and with global warming ; this question will be worked out with Doukhan, Heinen, Neumann, Renault, Truquet. See e.g. [KT2019].

Mostly the ecological data sets are mostly multivariate and they often admit very high dimensions. An important question is thus to select the coordinatess appropriate to deal with the question of interest. Renault plans to extend his previous work on causality measures in auto-regressive processes to the case of non-causal auto-regressive processes [GZ2017]. Such processes, when considered with heavy-tailed errors, have non-linear dynamics that allow for local explosion or asymmetric cycles.

Ecological ressources evolve considerably across the time and are influenced by many factors, climate characteristics among others. How to select the main relevant features for predicting or finding causal links for such quantities in the global warming period is a crucial problem, see [W1954].

Causal relations are needed in order to limit the dimension of the above vector valued time series, they also have a main importance for the explanation of ecological problems as the impact of salinity or nebulosity on the life of corals and on the marine ecosystem, with Doukhan, Garnier, Renault, and Navarrete.

More typical tools related to high dimension will also be used to work out information for high dimension time series ; beyond the classical sparsity assumptions, one may use low rank assumptions as in [ABDG2020] ; more general non linear models will be also exhibited in order to conveniently fit the huge sets of data considered in the above fields, see e.g. [FP2020] with the use of sparse assumptions.

Other issues of interest such as extreme values theory will also be considered for their econometrical and predictive applications, see [EKM2013].

Applications to the global warming, to the coral reefs data, and to marine resources will be developed together with the ecological team.

Task 5- **Epidemiology, Genomics and Biostatistics**

This task aims at study and predict epidemic episodes in the pandemic context of COVID19 and after it, too. In this frame, the importance of covariates is patent, and thus a direct relation together with tasks 3 and 4 is clear. The causality issues are essential in order to determine the impact of treatment on a given pathology.

In order to develop models adapted to epidemic behaviors we will introduce adapted integer valued models [DLN2020], possibly with negative values as in [DMN2020], in order to rate day-by-day evolutions. Biostatistics studies will be helpful for a deeper comprehension of the phenomenon. Their analysis may result in modeling of DNA, see [DFL2017]; Genetics data at summary level can be obtained from COVID-19 host genetics initiative (https://www.covid19hg.org/). Genomics variations of the SARS-CoV-2 itself are also available in GenBank, National MicrobiologyData Center and NGDC Genome Warehouse. Such data are high dimensional and dependent (in Linkage Disequilibrium), which related with task 4. Together those information may indeed be considered as covariates and/or causality for the study of epidemic behaviors.

It is also essential to conduct the economic issues of possible further pandemics in order to minimize risks in such occurrences.

Bibliography

[ABDG2020] Alquier P, Bertin, K Doukhan P, Garnier R (2020). High dimensional VAR with low rank transition. Statistics and Computing.

[ALT2010] Albrecher H, Ladoucette SA, Teugels JL (2010). Asymptotics of the sample coefficient of variation and the sample dispersion. Journal of Statistical Planning and Inference 140-2, 358-368.

[AT2006] Albrecher H, Teugels JL (2006). Asymptotic analysis of a measure of variation. Theory Probab. Math. Stat. 74, 1-9.

[BD2018] Bardet JM, Doukhan P (2018). Estimation of AR(1)-non stationary models. EJS 12, 2323-2354

[BKW2012] Bardet J-M, Kengne, W, Wintenberger O (2012) Detecting multiple change-points in general causal time series using penalized quasi-likelihood. Electronic Journal of Statistics 6, 435-477.

[BJ2007] Bascompte J, Jordano P (2007). Plant-animal mutualistic networks: the architecture of biodiversity. Annu. Rev. Ecol. Evol. Syst. 38, 567–593.

[BCD2019a] Bercu B, Capderou S, Durrieu G (2019) A nonparametric statistical procedure for the detection of marine pollution, J of Applied Statistics, 46(1), 119-140.

[BCD2019b] Bercu B, Capderou S, Durrieu G (2019) Nonparametric recursive estimation of the derivative of the regression function with application to sea shores water quality, SISP 22(1), 17-40.

[BSZ2006] Boxma O, Stadje W, Zacks S (2006). A Markovian growth-collapse model. AAP 38, 221-243.

[BCD2017] Brown M, Cohen JE, De La Pena V (2017). Taylor's laws, via ratio, for some distributions with infinite mean. J. Appl. Prob. 54, 657–669.

[CFM2006] Champagnat N, Ferrière R, Méléard S (2006). Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. Theoretical population biology 69-3, 297-321

[C1998] Cohen JE (1998). How Many People Can the Earth Support? Bulletin of the American Academy of Arts and Sciences 51-4, 25-39

[CMPR12] Cohen JE, Plank MJ, Law R. "Taylor's law and body size in exploited marine ecosystems." Ecology and evolution 2.12 (2012): 3168-3178.

[CX2015] Cohen JE, Xu M (2015). Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling. Proc. Nat. Acad. Sci. USA 112, 7749–7754.

[CBN2012] Cohen JE, Briand F, Newman CM (2012). Community food webs: data and theory. Springer Science & Business Media 20.

[DRW2019] Dahlhaus R, Richter S, Wu WB (2019).Towards a general theory for nonlinear locally stationary processes. Bernoulli 25- 2, 1013-1044

[DDLLLP2007] Dedecker J, Doukhan P, Lang G, León JR Louhichi S, Prieur C (2007). Weak Dependence: With Examples and Applications. LNS 190. Springer-Verlag, New York.

[D1994] Doukhan P (1994). Mixing, properties and examples. LNS 85. Springer-Verlag, New York.

[D2018] Doukhan P (2018). Stochastic Models for Time Series. Springer, Math and Applic 80.

[DFL2017] Doukhan P , Feugeas JP, Li X (2017) Statistical inference for DNA sequences of promoters, a non stationary qualitative model. Statistics 51-1, 154-166 (2017).

[DLN2020] Doukhan P, Leucht A, Neumann MH. Mixing properties of non-stationary INGARCH(1,1) processes. In preparation.

[DL1999] Doukhan P, Louhichi S (1999). A new weak dependence condition and applications to moment inequalities. Stochastic Processes and their Applications 84 (2), 313-342 441.

[DMN2020] N. Mamode Khan, M. H. Neumann. Mixing properties of Skellam-GARCH processes. https://arxiv.org/pdf/2005.12093

[E2018] C Emer et al. (2018). Seed-dispersal interactions in fragmented landscapes- a metanetwork approach. Ecology letters 21.4, 484- 493.

[DGPT2015] Durrieu G, Grama I, Pham QK, Tricot JM (2015) Nonparametric adaptive estimator of extreme conditional tail probabilities and quantiles, Extremes 18, 437-478.

[DPFM2016] Durrieu G, Pham QK, Foltete AS, Maxime V, Grama I, Le Tilly V, Duval H, Tricot JM, Sire O (2016) Dynamic extreme values modeling and monitoring by means of for sea shores water quality biomarkers and valvometry, Environmental Monitoring and Assessment, 188, 401-409.

[EKM2013] Embrechts P, Klüppelberg C, Mikosch T (2013). Modelling Extremal Events: for Insurance and Finance. Springer Science & Business Media.

[FLN2018] Fokianos K, Leucht A, Neumann MH (2018). On Integrated L^1 Convergence Rate of an Isotonic Regression Estimator for Multivariate Observations. arXiv:1710.04813v2.

[FNN2016] Fortuna L, Nunnari G, Nunnari S (2016). Nonlinear Modeling of Solar Radiation and Wind Speed Time Series, Springer

[GH2019] Goncalves B, Huillet T (2019). Scaling Features of Two Special Markov Chains Involving Total Disasters. Journal of Statistical Physics 1–33

[GX2017] Gouriéroux C, Zakoian JM (2017). Local explosion modelling by noncausal process. Journal of the Royal Statistical Society: Series B, 79, 737-756.

[HKS2018] Heinen A, Khadan J, Strobl E (2018). The Price Impact of Extreme Weather in Developing Countries. The Economic Journal.

[FP2020] Fermanian J-D, Poignard B (2020) High-dimensional penalized ARCH processes. Econometric Reviews

[K2015] S Kéfi et al. (2015). Network structure beyond food webs: mapping non- trophic and trophic interactions on Chilean rocky shores. Ecology 96.1, 291–303.

[KT2019] Fokianos K, Truquet T (2019). On categorical time series models with covariates. SPA 129-9, 3446–3462.

[M2014] PA Marquet et al. (2014). On Theory in Ecology. BioScience 64.8, 701-710.

[M2017] PA Marquet et al. (2017) On the proportional abundance of species: Integrating population genetics and community ecology. Nature Scientific Reports 7.1, 16815.

[M2018] P Marquet et al. (2018). Modelling the current and future biodiversity distribution in the Chilean Mediterranean hotspot. The role of protected areas network in a warmer future. Diversity and Distributions 25:1897–1909.

[M1992] Martinez ND (1992). Constant connectance in community food webs. The American Naturalist 139.6, 1208-1218.

[M1972] May RM (1972). Will a large complex system be stable? Nature 238.5364, 413 (1972).

[P-M2017] A Perez-Matus et al. (2017) Temperate rocky subtidal reef community reveals human impacts across the entire food web. Marine Ecology Progress Series 567, 1-16.

[R2019] R Rebolledo et al. (2019). An Open-System Approach to Complex Biological Networks. SIAM Journal on Applied Mathematics 79.2, 619-640.

[SR2002] Smith B, Rayens W (2002). Conditional Generalized Liouville Distributions on the Simplex. Statistics 36, 185-194.

[SR2002a] Smith B, Rayens W (2002). Multivariate Positive Dependence on the Simplex. Statistics 36, 75–88.

[W1954] Wold H (1954) Causality and Econometrics. Econometrica 22-2, 162-177

[Z2018] Zhang Z (2018) Multivariate Time Series Analysis in Climate and Environmental Research, Springer.