CIV-ML's Seminars 2017-2018

A General Hybrid Clustering Technique (Saeid Amiri, Post-Doc, Polytechnique Montreal, 29/08/2018)
Here, we propose a clustering technique for general clustering problems including those that have non-convex clusters. For a given desired number of clusters K, we use three stages to find a clustering. The first stage uses a hybrid clustering technique to produce a series of clusterings of various sizes (randomly selected). The second stage stabilizes the result of stage one by reclustering via the membership matrix under Hamming distance to generate a dendrogram. The third stage is to cut the dendrogram to get K_* clusters where $K_* > K$ and then prune back to K to give a final clustering. We provide a series of arguments to justify the steps in the stages of our methods and we provide examples involving real and simulated data to compare our technique with other techniques.

Analysis of concrete dam drainage flow data using BDLM (Simon Brousseau, Master student, Polytechnique Montreal, 10/08/2018)
Water infiltration through cracks and joints in concrete gravity dams can cause important deterioration to the structure. This defect is also hard to prevent and rectify. One way to deal with this issue is to collect the infiltration water across the structure and the monitor the flow in order to identify problematic zones. This seminar will present the analysis of the drainage water flow data of a concrete dam. The objective of the analysis is to predict the future response of the structure in regards of water infiltration.

Reinforcement Learning - Part 2: Model-free (Q-learning) (Shervin Khazaeli, Ph.D candidate, Polytechnique Montreal, 01/08/2018)
In previous seminar we covered the formalism of Reinforcement Learning (RL). The objective of RL is to map situations to actions without having a supervisor by maximizing the rewards. In this context RL can be divided in two categories namely model-based and model-free, where the model of the problem is respectively known and unknown. Here, the model is defined as the transition probabilities between different possible states of the system. However, in many real-world application, we do not have access to the model of the environment. Instead, we just experience the results of our actions. As such, the need for learning without knowing the model has been increased. In this seminar we will cover the basics of so-called Q-Learning, which is suit-

able for model-free applications. Our focus in this seminar lies in the context of infrastructure maintenance planning: Given the population of bridges, the objective is to make sequential decision (e.g. structural replacement) based on the state of the structure (e.g. poor condition).

Non-Periodic Kernel Regression for Time Series Filtering and Forecasting (James-A. Goulet and Luong Ha Nguyen, Professor and Ph.D. student, Polytechnique Montreal, 25/07/2018)
This seminar will present a new method called Non-Periodic Kernel Regression. This method builds on the theory of state-space models (SSM) for estimating hidden state variables. The method is intended to be employed in the context of time series displaying non-periodic patterns, for which no SSM-compatible method is currently available. The presentation will cover both the methodological aspects as well as applications of the method to case studies.

Using the Huang-Hilbert transform for choosing the model class in BDLMs (Ianis Gaudot, Post-Doc, Polytechnique Montreal, 18/07/2018)
Finding the right model class (type and number of components) in BDLMs is a critical issue. In most cases, the choice of the model class results from several test and trials, thus limiting the large-scale deployment of BDLMs. In this seminar, we propose to use the Huang-Hilbert transform to help choosing the right model class. The Huang transform decomposes a time series in a set of intrinsic mode functions (IMFs) along with a trend. The Hilbert transform allows to get the instantaneous attributes (amplitude and period) of each intrinsic mode functions. Here, we propose to model each IMF using a different BDLM periodic component. The period and process noise values of each BDLM periodic component are given from statistics on the instantaneous attributes computed from the Hilbert transform.

Reinforcement Learning - Part 1: Model-Based (Shervin Khazaeli, Ph.D candidate, Polytechnique Montreal, 10/07/2018)
Learning from interactions with environment is the basis of human learning process in which the learner acts on the environment and improves his/her action(s) by maximizing the corresponding reward(s). Inspired by such a trial-and-error process, Reinforcement Learning (RL) is introduced to the scientific society to build an agent that can map situations to actions without having a supervisor. This seminar is divided into two parts, namely

model-based and model-free, where the model of the problem is respectively known and unknown. During the first seminar we introduce the formalism of RL in the context of engineering and tackle the problem of infrastructure maintenance planning: Given the population of bridges, the objective is to make sequential decision (e.g. structural replacement) based on the state of the structure (e.g. poor condition). In the next seminar we will focus on the model-free RL, where the agent explores an unknown environment to perceive the optimized actions under different situation.

Kernel Regression approach for non-harmonic covariates estimation in Bayesian Dynamic Linear Model (BDLM) (Shervin Khazaeli, Ph.D candidate, Polytechnique Montreal, 13/06/2018)
Kernel approach is a powerful tool to extract patterns within timeseries. In this seminar we use static Kernel Regression (KR) to model Periodic (including harmonic and non-harmonic) in the context of Bayesian Dynamic Linear Models (BDLM). Data are simulated by newly developed âĂIJData LoaderâĂİ module considering baseline, seasonality, autoregressive, and measurement noise. The key aspects of this presentation are (i) examine the performance of the KR and (ii) Use correct definition (including seasonality, adaptive coefficient, and signal-to-noise ratio) for model construction and evaluation.

Using OpenBDLM - Part 1: Data loading (Ianis Gaudot, Post-Doc, Polytechnique Montreal, 23/05/2018)
The generalization of Bayesian Dynamic Linear Models (BDLMs) for Structural Health Monitoring (SHM) borrows tools from many fields (including applied statistics, machine learning, and signal processing), and there is currently no standalone software to allow for BDLMs to be used routinely by practioners. This presentation intends to bridge this gap by introducing OpenBDLM: a Matlab open source software specifically designed to use BDLMs for SHM. The software implements the four stages required for BDLM analysis, namely: (i) data loading and pre-processing, (ii) model construction, (iii) model parameters learning, and (iv) baseline response extraction. This seminar will present the features of the first stage that performs data loading and pre-processing.

Merging State-Space Models and Kernel Regression (James-A. Goulet, Professor, Polytechnique Montreal, 15/05/2018)

This seminar will present how the methods employed in Kernel Regression can be adapted to state-space model for learning patterns in time series. Three variants will be presented; two of them are suited for learning periodic patterns and a third one for non-periodic.

Metamodel-based Seismic Fragility Assessment of Concrete Gravity Dams (Rocio Lilen Segura, Sherbrooke University, 01/05/2018)
The consequences of a dam failure could be significant, both in terms of casualties and/or economic and environmental damage, for which its safety is given highest priority. Most gravity dams in Quebec were designed and built during the last century with methods of analysis which are considered inadequate today. With the advance of computational capabilities, numerical models have become more feasible and manageable, constituting the basis of more adequate procedures for designing and assessing. Seismic response and vulnerability assessment of dam-type structures, often requires a large number of non-linear dynamic analyses of complex finite element models to cover the predictor parameter space. The substantial computation time may be reduced by using statistical machine learning techniques to develop surrogate models. In addition given that earthquake shaking represents complex loading to a structure, it cannot be accurately characterized by a single intensity measure. In current vulnerability assessment methods, fragility functions are often developed using a single parameter to relate the level of shaking to the expected damage. This standard methodology neglects the variability in the estimated damage caused by the use of a single ground-motion parameter, which means that this uncertainty cannot properly be propagated to subsequent parts of the risk analysis nor can the importance of this variability be assessed. Addressing the existing disadvantages of unidimensional fragility functions and the lack of systematic exploration of different meta-modeling strategies to predict dam responses, this study analyzes the performance of several regression techniques. These meta-models are then used to develop multi-dimensional seismic fragility functions for the sliding limit state of gravity dams. The proposed methodology is applied to a case study dam, in north-eastern Canada. The seismic fragility assessment is complemented by considering a probabilistic based method to perform the record selection of accelerograms to be consistent with the probabilistic framework.An introduction of Stochastic Gradient Descent to Bayesian Dynamic Linear Models along with the illustrative examples will be presented in this seminar.

Stochastic Gradient Descent for Bayesian Dynamic Linear Models (Luong Ha Nguyen, Ph.D student, Polytechnique Montreal, 25/04/2018)
An introduction of Stochastic Gradient Descent to Bayesian Dynamic Linear Models along with the illustrative examples will be presented in this seminar.

Why does Kernel approach work ? (Shervin Khazaeli, Ph.D candidate, Polytechnique Montreal, 18/04/2018)
In this seminar, the Kernel approach is presented from the classical point of view and modern one in Machine Learning (ML). It is tried to introduce different perspective towards such approach for particular problems. An interestingly question here is what is Kernel and above that why does it work astonishing in many disciplines. Although, the concept of Kernel is defined in various fields of study such as modern algebra, machine learning, statistics, and so on, here, we focus on the statistic and ML points of view. Finally, the application of utilizing Kernels in Bayesian Dynamic Linear Models (BDLM) is investigated for recognizing different patterns in time series.

Statistical downscaling of daily and hourly climate scenarios for the various meteorological variables at Montreal (Barghob Deka, M. Eng., McGill University, 11/04/2018)
Environmental assessment studies for the future decade under the influence of a changing climate is the need of the hour. In order to identify the future synoptic weather types, future daily and hourly projections of the weather variables are required. Hence, this research is motivated by the lack of a comprehensive and statistically significant downscaling methodology for the various weather variables. The present study is based on the various statistical downscaling techniques developed by the researchers in the past using linear regression model because of the advantage of being less computationally intensive. In addition, an attempt has been made to develop an improved statistical downscaling methodology by combining different techniques to develop a robust method with a detailed performance assessment of the models. Linear Regression models are derived to downscale daily climate scenarios using NCEP reanalysis datasets for the predictors and weather station data available at Trudeau International Airport for the predictands during the reference period 1958-2001. The NCEP predictors are regridded to the GCM scale as the GCM outputs are to be used for deriving future climate projections. Standardization and Deseasonalization of the predictor variables are carried out followed by principal component analysis prior to their introduc-

tion into the stepwise regression model. CANESM2 is selected as the GCM model in the present work whose outputs are used as predictors in the NCEP derived regression models. A Bias Correction procedure is used to correct the systematic biases present using a quantile-quantile mapping technique on the downscaled variable using CANESM2 predictors. After downscaling the daily climate variables, hourly downscaling transfer functions are derived based on the historical relationships of the hourly values with its daily mean as well as other weather predictors where appropriate. For the future climate projections, RCP2.6, 4.5 and 8.5 are used as greenhouse gas trajectories representing the change in climate in the future decade.

Criteria for model class verification in Bayesian dynamic linear models (Ianis Gaudot, Post-Doc, Polytechnique Montreal, 03/04/2018)
Given a pre-defined model class and a set of optimized model parameters, the model class verification aims at evaluating if the results behave as expected. In this seminar, we propose some criteria based on the residual component to perform model class verification in BDLM.

Tutorial - Graphical User Interface (GUI) design in MATLAB (Zachary Hamida, Ph.D candidate, Polytechnique Montreal, 27/03/2018)
This tutorial will cover general concepts about GUI design in MATLAB. The session will include handling objects and functions in GUI environment and creating a ready to publish GUI.

Environmental effects with PCA in BDLM (Catherine Paquin, Master student, Polytechnique Montreal, 13/03/2018)
In this seminar, a method using PCA decomposition in order to include correlated data in BDLM is presented. The detailed method and an application with real data are explained and analyzed.

Validating the Accuracy of Forecast Models in the Case of Limited Historical Data (Zachary Hamida, Ph.D candidate, Polytechnique Montreal, 21/02/2018)
The performance of forecast models is commonly evaluated through different validation methods. The efficiency of these methods is mainly dependent on the amount of data available. Therefore, in the cases where the data is limited (small datasets), the regular validation methods becomes insufficient. In this seminar, a framework for validating forecast models in small datasets

will be presented. This framework will be employed in evaluating the performance of different forecast models for visual inspection data.

Uncertainty of model parameters and hidden states in Bayesian Dynamic Linear Models (Luong Ha Nguyen, Ph.D student, Polytechnique Montreal, 07/02/2017)
The presentation introduces two procedures; (1) the Maximum a Posteriori with Laplace approximation procedures and (2) the Hamilton Monte Carlo procedure for uncertainty quantification task in the existing Bayesian Dynamic Linear Models. A comparative study between both procedures is conducted on the simulated data and the real data collected on a dam in Canada.

A comparison between Hydrostatic-Time-Season and Bayesian Dynamic Linear models for monitoring behavior of dams (Ianis Gaudot, Post-Doc, Polytechnique Montreal, 31/01/2018)
In the field of dam monitoring, the most popular data interpretation method is the Hydrostatic-Season-Time (HST) model, which performs a regression analysis to decompose a time series into sub-components. Despite its widespread application, HST models suffer from two major drawbacks, (i) they are not suited to process non-stationary time series, and (ii) anomaly detection is based on hypothesis testing procedure. On the other hand, Bayesian Dynamic Linear Models (BDLM) are able to overcome these limitations. This seminar will discuss the advantages and limitations of the two approaches based on simulated and real data.Particle filtering is a sequential Monte-Carlo approach which is well suited to handle nonlinear dynamic systems. At each iteration, a set of Monte-Carlo samples (particles) must be sampled from a distribution called the importance distribution. The choice of the importance distribution has a strong impact on the performance of the method. In this presentation, we will review some popular approaches found in the literature. The performance of each algorithm will be illustrated on simple examples in the context of structural health monitoring.

Generalization of Hidden non-harmonic covariates estimation by Non Uniform Rational Basis Splines (NURBS), Part 1: Fundamentals and Data Simulation (Shervin Khazaeli, Ph.D candidate, Polytechnique Montreal, 24/01/2018)
Many structural responses are affected by hidden non-harmonic covariates such as temperature and traffic loadings, to name a few. In Bayesian Dy-

namic Linear Models (BDLM) we may include such effects by employing Dynamic Regression Component (DRC). In new the approach, one can replace the observation matrix in DRC by a hidden response function that mimics the behavior of the hidden covariate. In this seminar it is tried to utilize Non Uniform Rational Basis Splines (NURBS) approach to generate any flexible function with respect to continuity and orders for above-mentioned purpose. To this end, NURBS fundamentals are investigated and then the algorithm is presented to simulate synthetic data considering such effect during the monitoring (observation).

Handling Constraints in Kalman Filter Framework (Zachary Hamida, Ph.D candidate, Polytechnique Montreal, 17/01/2018)
Constraints are mathematical formulas that assure the parameters being updated (or optimized) stay within realistic bounds. The application of constraints is essential in any pure optimization problem. However, constraints are rarely mentioned in regression problems, specifically, time series forecast problems. In this seminar, the use of constraints in time series problems will be addressed along with some methodologies for applying constraints within (the popular) Kalman Filter. Moreover, real-life application and examples will be presented from the MTQ dataset.

Environmental effects with PCA in BDLM - A case study (Catherine Paquin, Master student, Polytechnique Montreal, 29/11/2017)
This seminar presents the methodology developped to include the environmental effects in the analysis of a structureâĂŹs behavior using PCA. A case study is detailed to demonstrate the application of the method, and the results are analyzed.

An overview of Hamiltonian Monte Carlo (Luong Ha Nguyen, Ph.D student, Polytechnique Montreal, 22/11/2017)
The presentation introduces the Hamiltonian Monte Carlo (HMC) to the Bayesian estimation. An introduction about the theory behind HMC along with the typical examples will be provided in this presentation.

On the Influence of the Importance Distribution Choice in Particle Filtering (Ianis Gaudot, Post-Doc, Polytechnique Montreal, 15/11/2017)
Particle filtering is a sequential Monte-Carlo approach which is well suited to handle nonlinear dynamic systems. At each iteration, a set of Monte-Carlo

samples (particles) must be sampled from a distribution called the importance distribution. The choice of the importance distribution has a strong impact on the performance of the method. In this presentation, we will review some popular approaches found in the literature. The performance of each algorithm will be illustrated on simple examples in the context of structural health monitoring.

High-Dimensional Data Visualization Using t-SNE Technique (Shervin Khazaeli, Ph.D candidate, Polytechnique Montreal, 08/11/2017)
In the context of Structural Health Monitoring (SHM), data-driven models often utilize high-dimensional space due to numerous variables. This presentation is dedicated casting data visualization in high-dimensional data space using student t-distributed Stochastic Neighbor Embedding (t-SNE). The objective of the technique is to give each datapoint in high-dimensional space a location on two or three dimensional space preserving the global and local structure. First, the techniques is described in details by providing simple case example. Afterwards, neat examples are investigated to reveal the capability of such technique in SHM.

Improving Prediction Models Capacity in Time Series Problems (Zachary Hamida, Ph.D candidate, Polytechnique Montreal, 01/11/2017)
Real-life data commonly consume high nonlinearity due to erratic measurements or due to neglecting some important factors. Therefore, when attempting to model and extract information from real-life data (i.e. perform predictions), ordinary models may perform poorly. Boosting is one of the techniques used to overcome this limitation. In this seminar, a boosted prediction framework will be presented with application on MTQ bridges dataset. The objective of the prediction model will be to predict the state of a bridge beam elements over time.

Gaussian Process Regression with Categorical Covariates (James-A. Goulet, Professor, Polytechnique Montreal, 18/10/2017)
Gaussian Process Regression (GPR) is an established method for building empirical models relating continuous covariates and observed system responses. There are few methods available for employing GPR with categorical covariates describing non-ordered categories. e.g. bus, car, taxi. This presentation will expose the limitations of existing methods and propose a new methodology for overcoming these limitations.

Principal Component Analysis in Bayesian Dynamic Linear Models (Catherine Paquin, Master student, Polytechnique Montreal, 04/10/2017)
The use of more than one sensor to measure the same physical quantity has many advantages including reliability and added information. The principal component analysis is used to include the information from many sensors in a BDLM analysis. The gain of precision in the different cases are then compared based on the log-likelihood of each combination. The data used for the example are from an existing highway overpass.

Multi-pass and Dropout with Bayesian Dynamic Linear Models for Anomaly Detection (Luong Ha Nguyen, Ph.D student, Polytechnique Montreal, 27/09/2017)
The false alarm is the main factor limiting the applicability of the Online Batch Procedure in Bayesian Dynamic Linear Models. This presentation introduces the Multi-pass and Dropout techniques for addressing this challenge. The potential of the new techniques is illustrated on displacement data recorded on a dam in Canada.

An Overview on SHM Dataset of Bridges in Quebec with Preliminary Analysis (Zachary Hamida, Ph.D candidate, Polytechnique Montreal, 20/09/2017)
This seminar will be delivered in two sections: the first section will introduce and examine the information available in the SHM dataset of bridges in Quebec. The dataset is currently being studied to analyze the degradation of infrastructures on a network scale. The second section of the seminar will present the steps towards using the aforementioned dataset in predicting future scenarios (i.e. degradation degree over time in structural elements). This will include handling biases and erratic measurements encountered naturally in real-life datasets.

Anomaly detection from the residual component in Bayesian dynamic linear models (Ianis Gaudot, Post-Doc, Polytechnique Montreal, 13/09/2017)
In Bayesian dynamic linear models, time series are decomposed into subcomponents. The type (local level, trend, acceleration, periodic) and the number of the sub-components are chosen a priori by the user. In real-life applications, it is a common practice to add a residual component in order

to account for missing effects which can not be known a priori. The residual component is generally modeled by a first-order autoregressive model (AR(1)). During this presentation, we will investigate the possibility to detect anomaly directly from this residual component. The method is based on regime-switching models and the analytic formulation of AR(1) models. The advantages and limitations of this technique with respect to classical approaches used for anomaly detection (which usually exclude residual component) will be discussed.