LEARNING BAGGED MODELS OF DYNAMIC SYSTEMS

NIKOLA SIMIDJIEVSKI[1,2] , LJUPCO TODOROVSKI[3] , SAŠO DŽEROSKI[1,2]
1 DEPARTMENT OF KNOWLEDGE TECHNOLOGIES, JOZEF STEFAN INSTITUTE, LJUBLJANA, SLOVENIA
2 JOZEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL, LJUBLJANA, SLOVENIA
3 FACULTY OF ADMINISTRATION, UNIVERSITY OF LJUBLJANA, SLOVENIA
NIKOLA.SIMIDJIEVSKI@IJS.SI

INTRODUCTION

*Equation Discovery (ED) is the process of inducing scientific laws and models in form of equations from observations (measured data)

*Inductive Process Modeling (IPM) based on ED uses combination of domain knowledge and observations (measurements) to induce explanatory process-based models of dynamic systems

*Ensemble learning methods (ENSEMBLES) consider combinations of multiple predictive models to maximize the predictive performance of the system

MOTIVATION

*Ensembles of dynamic models has not yet been considered in the machine learning community.

*Adapting the traditional Bagging approach, in the context of learning ensembles of ODE models

AIM : Evaluate the performance of ensemble of ODE models compared to classical ODE models with automated process based modeling approach

ProBMoT (Čerepnalkoski et. al.)

*ProBMoT - tool for complete modeling, parameter estimation and simulation of process-based models

*Conceptual (high-level) representation of a model
  - Rules which indicate how the model structures are generated

*Models are formulated as ODEs
*Knowledge is represented as lbrary of entities and processes
*Entities are state/variable of the system
*Processes represent the interaction between the entities

*Parameter estimation is based on meta-heuristic optimization
*ProBMoT implements custom quantitative objective functions
*Modification of ProBMoT to meet the criteria for generating ensembles of ODE models

ENSEMBLES OF ODE models

Our approach chooses uniformly random time points from the time series & assigns weights to each of them.

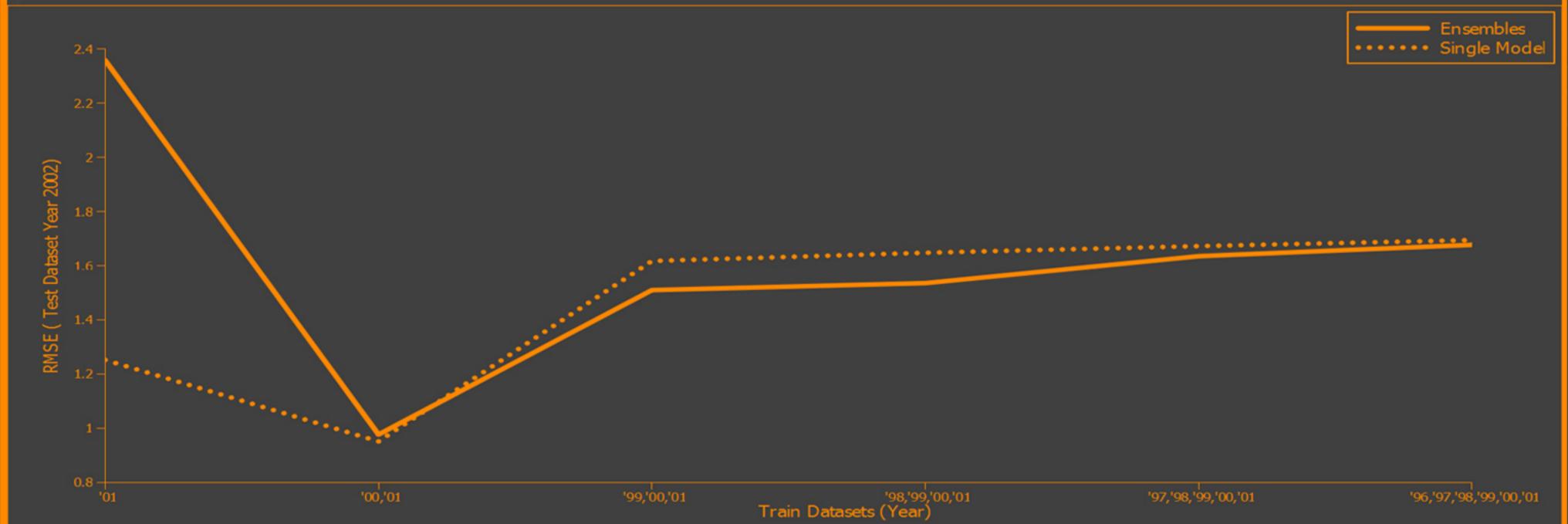By penalizing (encouraging) different time points from each boot-

Adapted bagging

EXPERIMENTS & CONCLUSION

Lake Bled , Slovenia
Modeling phytoplanckton concentration
Train Set : 6 datasets containing measurements by year (1996-2001)
Test Set : 1 dataset with measurements of year 2002
Adapted Bagging for ODE models
Evaluation scenario : Learning Curve

The ensemble model only slightly improves the overall performance of the ODE models
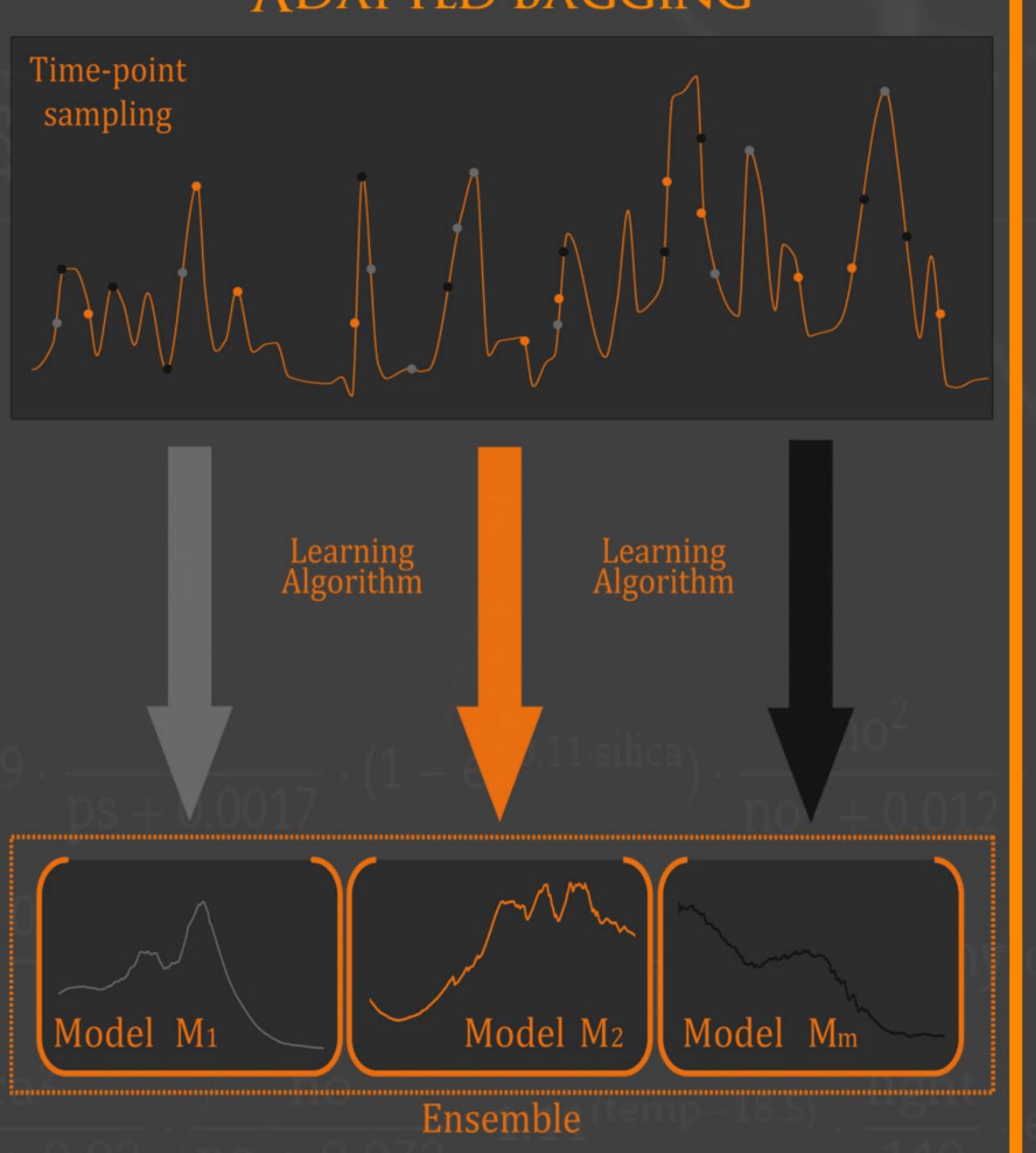
Adding too many preceding data sets to the training data doesnt improve the predictive performance

These results uncover important characteristics of the ensembles approach. Moreover, they clearly point to important directions for future improvement of the performance of this methodology

FOOD-WEB DYNAMICS IN AQUATIC ECOSYSTEMS

EXAMPLE MODEL

growth
$$td(phyto.conc) = phyto.maxGrowthRate * phyto.conc * phyto.nutrientLim$$

$$td(nitrogen.conc) = -nitrogen.alpha * phyto.maxGrowthRate * phyto.conc * phyto.nutrientLim$$

$$td(phosphorus.conc) = -phosphorus.alpha * phyto.maxGrowthRate * phyto.conc * phyto.nutrientLim$$

nitrogenLim
$$phyto.nutrientLim = nitrogen.conc / (nitrogen.conc + nitrogen.halfSat)$$

phosphorusLim
$$phyto.nutrientLim = phosphorus.conc / (phosphorus.conc + phosphorus.halfSat)$$

phyto
conc
nutrientLim
maxGrowthRate

nitrogen
conc
halfSat
alpha

phosphorus
conc
halfSat
alpha

CONCEPTUAL MODEL

Zooplankton
Grazing
Nutrient
Respiration
Growth
Primary Producer
Mortality
Sedimentation
Sediment