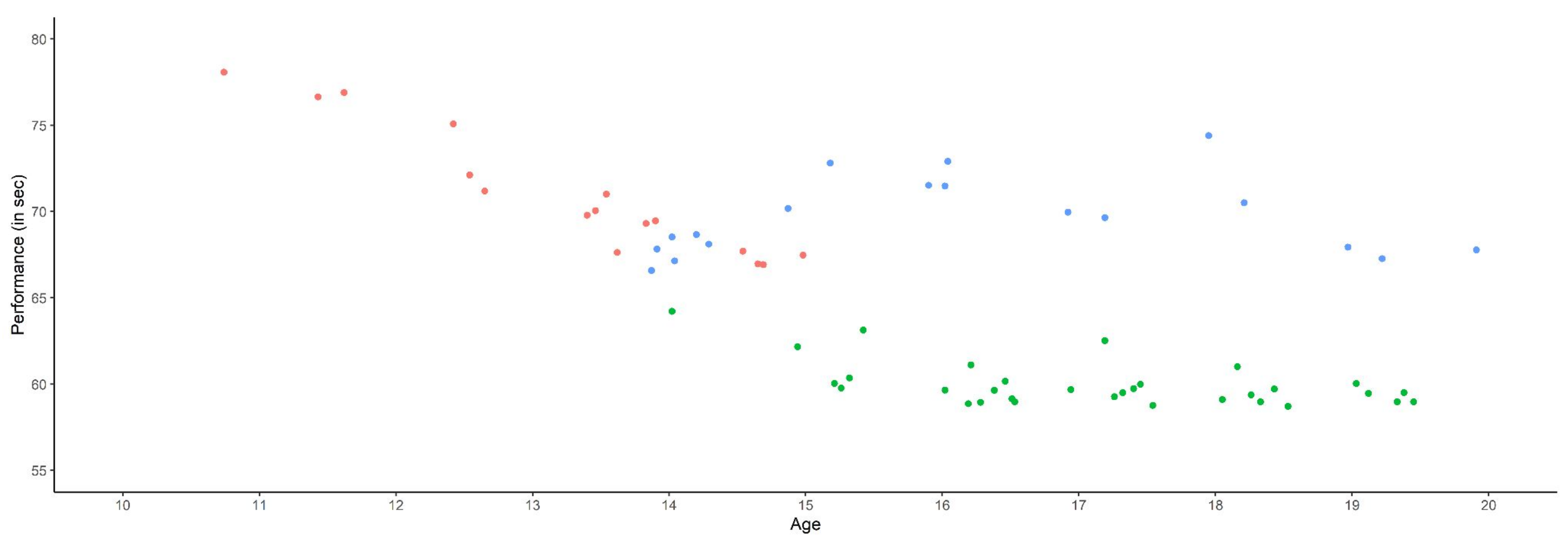
Cluster-Specific Predictions with Multi-Task Gaussian Processes

Arthur Leroy
Pierre Latouche
Benjamin Guedj
Servane Gey

Context

Suppose you observe irregular longitudinal data coming from multiple sources, collected at different locations (e.g. on the figure below, the performance on 100m freestyle events for 3 different swimmers have been observed at different ages). Multi-task (or multi-output) Gaussian processes (GPs) approaches have been successfully applied despite scaling and flexibility limitations [1,2,3].



We recently proposed a novel paradigm for defining multi-task GPs by **transferring knowledge through a common latent mean process** [4]. This framework has first been extended to perform a **simultaneous clustering and prediction** of the tasks in a non-parametric and probabilistic design [5].

Sharing information through mean processes

If longitudinal data are collected from M correlated tasks, let us assume that all the observations $y_i(t), i = 1, \dots, M$ can be allocated into K clusters and decomposed as the sum of a cluster-specific mean process μ_k, a task-specific process f_i and a noise term. Formally, if the i-th task belongs to the k-th cluster:

$$y_i(t) = \mu_k(t) + f_i(t) + \varepsilon_i(t),$$

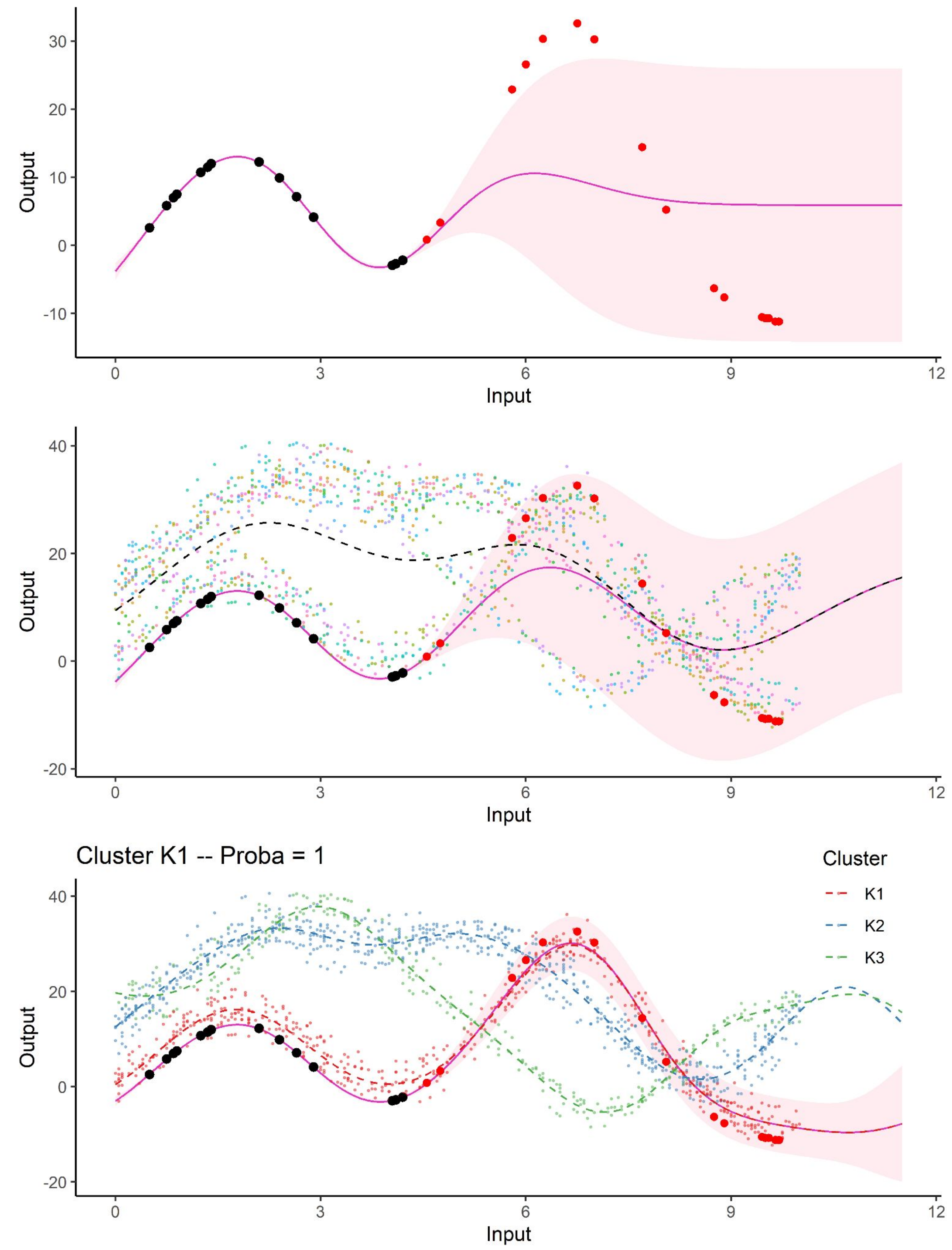
with

$$\mu_k \sim \mathcal{GP}(m_k, C_k), \quad f_i \sim \mathcal{GP}(0, \Sigma_i), \quad \varepsilon_i \sim \mathcal{GP}(0, \sigma_i^2 I) \quad \mathcal{Z}_i \sim \mathcal{M}(1, \pi)$$

where we do not impose any assumption on the covariance structures. Conditionally to the mean processes, all tasks are independent. Thanks to a **Variational EM algorithm**, we can still learn kernel hyper-parameters through gradient-descent optimisation (M step) and compute variational hyper-posterior approximations of the latent mean process in closed form (VE step). The learning procedure **complexity is linear** in the number of tasks/clusters. After learning, the key idea for prediction is to integrate out a mean process and define a **multi-task prior** distribution for any new task $\mathbf{y}_*$ and all clusters, as:

$$\begin{aligned} p(\mathbf{y}_* \mid \{\mathbf{y}_i\}_{i=1,\dots,M}) &= \int p(\mathbf{y}_*, \boldsymbol{\mu}_0 \mid \{\mathbf{y}_i\}_{i=1,\dots,M}) \mathrm{d}\boldsymbol{\mu}_0 \\ &= \int p(\mathbf{y}_* \mid \boldsymbol{\mu}_0) p(\boldsymbol{\mu}_0 \mid \{\mathbf{y}_i\}_{i=1,\dots,M}) \mathrm{d}\boldsymbol{\mu}_0 \\ &= \mathcal{N}(\mathbf{y}_*;\ \hat{m}, \hat{K} + \boldsymbol{\Sigma}_* + \sigma_*^2 I). \end{aligned}$$

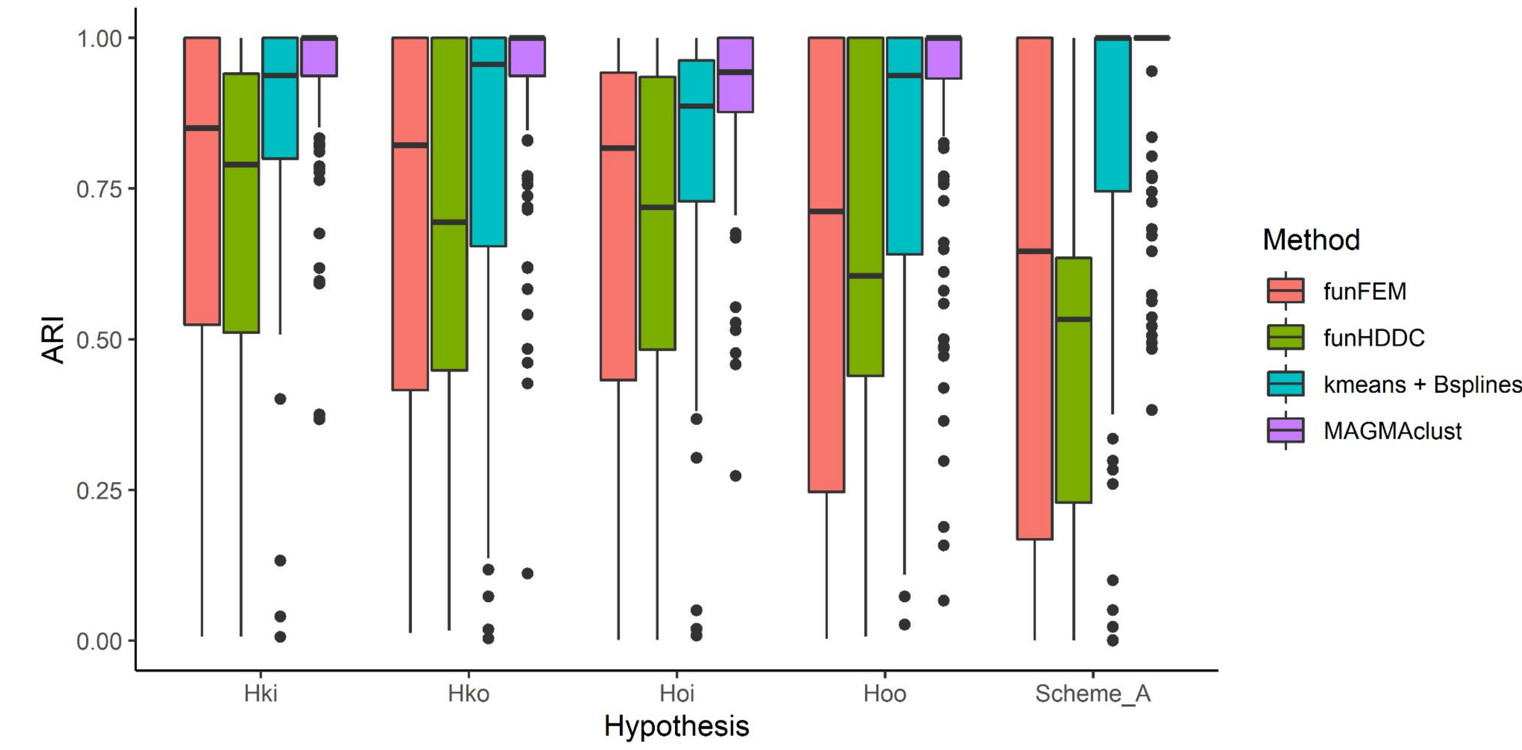
The R package ***MagmaClustR*** (https://arthurleroy.github.io/MagmaClustR) offers all tools to perform learning, prediction and visualisation within this framework.



In the figure above, we display a comparison between **standard GP** (top), the approach proposed in [4] called **Magma** (middle), and our extension named **MagmaClust** (bottom), to perform prediction on (red) testing points using (black) training points and information from other tasks (backward points). The training individuals are coloured according to the most probable cluster (bottom panel).

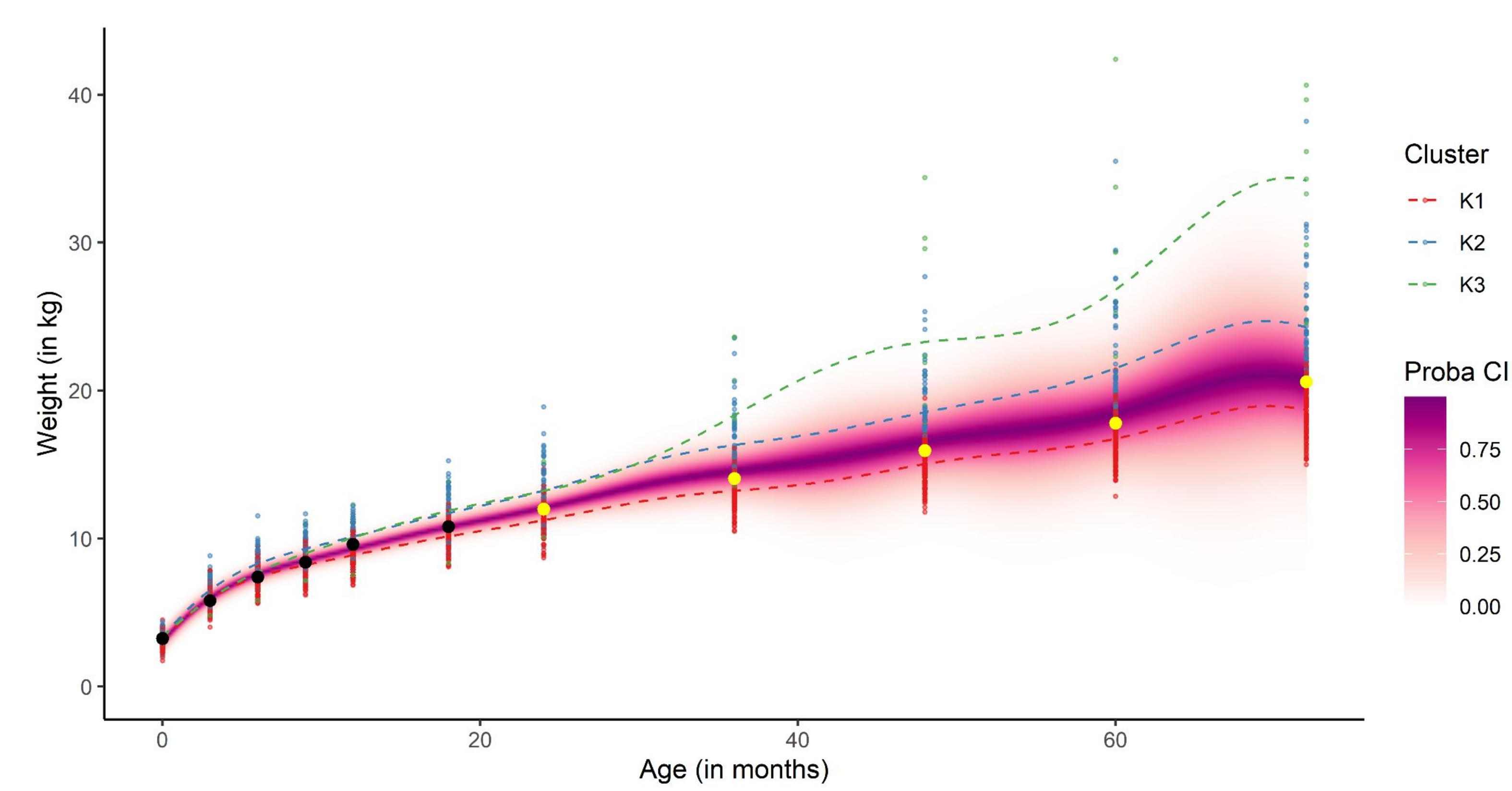
The ability to recover the underlying cluster of similar patterns from a handful of data points of a specific task improves dramatically the **quality of predictions** and **reduces the associated uncertainty**, by favouring the relevant information.

In the following figure, we also highlight the remarkable performances in terms of **clustering accuracy**, outperforming state-of-the-art curve clustering methods when computing the Adjusted Rand Index (ARI) for various simulation schemes.



Application to diverse problems

In the article, the MagmaClust algorithm has been successfully applied to a forecasting problem of **talent detection** for young swimmers, missing data **reconstruction of CO2 emission** time series, and **weight follow-up** during childhood (see figure below). Any problems involving correlated longitudinal, spatial or any continuous measurements is generally susceptible to be well handled by this approach, and the current implementation is designed to make this process as easy as possible with a few lines of code. **Try it, you will like it!**



References

[1] P. Goovaerts (1997) Geostatistics for Natural Resources Evaluation. In: Oxford University Press

[2] E. V. Bonilla et al. (2008) Multi-task Gaussian Process Prediction. In: Advances in Neural Information Processing Systems

[3] M. A. Álvarez et al. (2012) Kernels for Vector-Valued Functions: A Review. In: Foundations and Trends in Machine Learning

[4] A. Leroy et al. (2022) MAGMA: Inference and prediction using multi-task Gaussian processes with common mean. In: Machine Learning

[5] A. Leroy et al. (2022) Cluster-Specific Predictions with Multi-Task Gaussian Processes. In: Journal of Machine Learning and Research