Rapidly Inferring Personalized Neurostimulation Parameters with Meta-Learning: A Case Study of Individualized Fiber Recruitment in Vagus Nerve Stimulation

Ximeng Mao[1,2], **Yao-Chuan Chang**[3], **Stavros Zanos**[3], **Guillaume Lajoie**[1,4,5,*]

[1] Mila-Quebec Artificial Intelligence Institute, 6666 St-Urbain, Montréal, QC, H2S 3H1, Canada
[2] Department of Computer Science and Operations Research, University of Montréal, 2920 chemin de la Tour, Montréal, QC, H3T 1J4, Canada
[3] Institute of Bioelectronic Medicine, Feinstein Institutes for Medical Research, Manhasset, NY, 11030, USA
[4] Department of Mathmatics and Statistics, University of Montréal, 2920 chemin de la Tour, Montréal, QC, H3T 1J4, Canada
[5] Canada CIFAR AI Chair
[*] Corresponding Author

E-mail: `g.lajoie@umontreal.ca (G. Lajoie)`

Abstract. *Objective.* Neurostimulation is emerging as treatment for several diseases of the brain and peripheral organs. Due to variability arising from placement of stimulation devices, underlying neuroanatomy and physiological responses to stimulation, it is essential that neurostimulation protocols are personalized to maximize efficacy and safety. Building such personalized protocols would benefit from accumulated information in increasingly large datasets of other individuals' responses. *Approach.* To address that need, we propose a meta-learning family of algorithms to conduct few-shot optimization of key fitting parameters of physiological responses in new individuals. While our method is agnostic to neurostimulation modality, here we demonstrate its effectiveness on the problem of physiological modeling of fiber recruitment during vagus nerve stimulation (VNS). Using data from acute VNS experiments, the mapping between amplitudes of stimulus-evoked compound action potentials (eCAPs) and physiological responses, such as heart rate and breathing interval modulation, is inferred. *Main results.* Using additional synthetic data sets to complement experimental results, we demonstrate that our meta-learning framework is capable of directly modeling the physiology-eCAP relationship for individual subjects with much fewer individually queried data points than standard methods. *Significance.* Our meta-learning framework is general and can be adapted to any input-response neurostimulation mapping problem. Moreover, this method leverages information from growing data sets of past patients, as a treatment is deployed. It can also be combined with several model types, including regression, Gaussian processes with Bayesian optimization, and beyond.

Keywords: Learning of Personalized Parameters, Few-shot Learning, Meta-learning, Vagus Nerve Stimulation, Gaussian Process, Bayesian Optimization

Table 1: Table of Key Acronyms

Terminology	Acronym
Vagus Nerve Stimulation	VNS
Bayesian Optimization	BO
Gaussian Process	GP
(relative) Rooted Mean Square Error	(r)RMSE
Electromyograph	EMG
Heart Rate	HR
Breathe Interval	BI
Dataset Name	**Acronym**
synthetic exponential function	syn-exp
between EMG and A fiber recruitment	EMG-A
between changes in HR and B fiber recruitment	cHR-B
between changes in BI and C fiber recruitment	cBI-C
expanded synthetic data from cBI-C	semi-C-syn
Method Name	**Acronym**
meta-learning framework with GP as population model	metaGP
meta-learning framework with regression as population model	metaRG
metaGP with enabled inner-loop gradient update	metaGP-gu
metaRG with enabled inner-loop gradient update	metaRG-gu
GP only	baseGP
regression only	baseRG

1. Introduction

Neurostimulation is growing in popularity as a potential treatment for several ailments, for example, spinal cord stimulation (SCS) for chronic pain [1], deep brain stimulation (DBS) for Parkinson's disease [2] and vagus nerve stimulation (VNS) for epilepsy [3]. It is administered via either invasive or non-invasive approaches to deliver electrical stimulation to nervous system [4]. This can be done either in the central or peripheral nervous systems for varied targeted outcomes. An important part of developing neurostimulation treatments is to accurately estimate a subject's physiological response map to stimulation, which can then be used for personalized optimization. This is challenging even in the presence of ever growing datasets from past patients, mainly because of inherent variability in subject to subject responses, but also because of varying implant placement and other physiological factors [6]. It is therefore important to leverage this acquired data while avoiding the pitfalls of simply fitting and using population mean response models. To address this problem, we propose a flexible meta-learning framework to perform *few-shots* learning of key response fitting parameters given a newly implanted subject and a pool of past model fits data.

Few-shot learning [7,8] is a widely studied learning paradigm in machine learning, where the learning model is asked to solve each task while observing only a small fractions of the data. In the context of neurostimulation, each shot can be a query to the subject's response map. Therefore, the ability to do well in few-shot learning is beneficial for neurostimulation, considering both that it can reduce the number of stimulations that the patients take, and that it can avoid the amount of time required to process the additional signals. Attempting to generalize well with only a small training set can be challenging, but it can be facilitated by leveraging the experience of learning other tasks that are similar, which is the core concept of meta-learning [9, 10], also known as "learning to learn". The overarching training objective of few-shot learning, solved by meta-learning approaches, focuses on improving a model's ability to adapt and generalize with access to only few samples from each task, while leveraging meta learnt prior knowledge from previously available tasks. Note that in meta-learning, we have data both from other tasks as well as the few-shots from the current task. Therefore, to better distinguish between the two sources of information, the former is commonly termed as the meta-training set.

By viewing the personalizaion of one single subject as one task, we can draw direct analogy between it and meta-learning. The goal of our proposed meta-learning solution is to extract, from a population of past subjects, as much information that is useful for a personalized model to be optimized in open or closed loops settings. In this manuscript, we will concentrate on the physiological modeling of individual fiber recruitment in VNS and associated physiological responses such as heart rate and breathing interval modulation. We conduct our analysis based on the published data found in [20]. Importantly, we reiterate that our proposed framework is general and can be exploited in various modalities, as we discuss later on.

VNS is an emerging treatment for a multitude of disorders [3, 5, 12–16] including epilepsy, depression, anxiety, Alzheimer's disease, chronic pain and rheumatoid arthritis. It stimulates the vagus nerve, a main nerve of peripheral nervous systems that extends from the brainstem to the abdomen and is responsible for vast essential body functions, such as heart rate, breathing and digestion [11]. Accumulating evidence suggests that varying the stimulation parameters for individual subjects can improve the efficacy of VNS [17, 18]. Hence, to develop targeted and personalized clinical interventions, it is crucial to acquire quantifiable fiber recruitment during VNS on a single subject basis. Obviously, this requirement is shared with several other neurostimulation modalities, which makes the particular problem a representative case study to evaluate meta-learning.

Based on the known correlations between certain physiological markers and specific fiber groups [21, 22], one way to obtain the fiber amplitude is through modeling the physiological effects in response to the neurostimulation. Recent work by Chang *et al* [20] demonstrated linear or non-linear relationships in a rat model for each of the three pairs of physiological markers and fiber amplitudes of stimulus-evoked compound action potentials (eCAPs): between cervical electromyography (EMG) and A-fiber, between changes in heart rate (ΔHR) and B-fiber and between changes in breathing interval (ΔBI) and C-fiber. Moreover, they derived and validated regression models for all three pairs across all subjects, given each subject's normalized measurements, on responses of both physiological and fiber, over their respective maximal values. However, due to the requirements of a subject-dependent normalization factors, these regression models cannot be applied directly to unseen subjects, a requirement for these response models to be useful in deployed VNS treatments.

In this study, we apply our meta-learning framework to this problem, where the individual subject's operating ranges of both physiological and fiber responses' amplitudes are the key response fitting parameters to be learnt. More specifically, the framework is tasked with both the extraction of

prior knowledge from a pool of past subjects and the inference of the factors of normalization given the few shots from a new subject. Combining the two pieces of information, a personalized prediction model can then be derived for the new individual, using as few query points or "shots" (stimulation with distinct parameter values) as possible. With prior knowledge learnt from available meta-training data, our framework creates an average model (hereon referred to as population model), and then adapts it accordingly for each new subject given few shots of the measurements. The framework is by design compatible with any model type as population model, creating a family of algorithms, and we investigated two of them in our implementations: close-form regression and Gaussian process (GP) [27]. We validated the performance of the proposed meta-learning framework in multiple evaluations on both publicly available experimental data and synthetically created artificial data, and we found accurate and consistent performance across all evaluations regardless of the choice of population model.

Our main contribution is to propose meta-learning as a promising and generic solution to the challenge of personalization in neurostimulation, and as a concrete case study in physiological modeling of fiber recruitment in VNS, we present a meta-learning framework to infer individual subjects' models using only few data points. We provide evaluations on both synthetic and publicly available experimental data, to verify the few-shot prediction performance of the proposed meta-learning framework, especially when the number of few-shot is small.

The manuscript is structured as follows. In Section 2, we illustrate the process of utilizing the meta-learning framework to predict individual subject's fiber amplitudes, and we also show the results of evaluations on three different datasets, both real and synthetic ones, for algorithm variations using the framework. We also provide a discussion where we outline the limitation and future work in Section 3. Finally in Section 4, we start from the problem formulation of few-shot learning. as well as an overview of related works on both personalized treatment in neurostimulation and meta-learning, and then provide detailed descriptions on relevant algorithms involved in the proposed meta-learning framework, as well as the synthetic subject generation procedures.

2. Results

2.1. Meta-Learning Framework for Few-shot Prediction of Individual Subjects' Fiber Amplitudes

In Figure 1(a), we provide an illustration on how to use the proposed meta-learning framework to conduct few-shot prediction on individual subjects' fiber amplitudes, and the diagram of the framework is shown in Figure 1(b). Pseudo-code of the framework is presented in Algorithm 1. Acting as the backbone for all the future individual models, the framework compiles the population model a priori, using the normalized data from all subjects in the meta-training set, which in our case is more fitting to refer as the population set. Note that we will use meta-training set and population set interchangeably for the rest of this manuscript. The population model is supposed to reflect the prior knowledge extracted from the previous subjects, embodied here as a population mean fit. There is no restriction on the model types used as population model, with notable candidates as GP and closed-form regression. For readers familiar with meta-learning, the parameters of the population model can be viewed as the meta parameters, considering they are learned from the pool of population and serve as the extracted prior knowledge.

Algorithm 1 Few-shot Prediction on Individual Fiber Amplitudes from Physiological Responses

Require: population set D_{ppl} composed of data from n_s different subjects: $\{x_i^j, y_i^j\}$
Require: data from new subject k: $\{x_i^k, y_i^k\}$
Require: learning rate α and optionally β

Meta-Training

Fit population model $f_{ppl}(\cdot;\theta)$ on normalized D_{ppl}
while not done **do**
  for j=1,...,n_s **do**
    Infer $\mathbf{Z}^j = (Z_x^j, Z_y^j)$ by *Bayesian Optimization* taking N shots from subject j
    $f_j(\cdot;\phi,\mathbf{Z}^j) \leftarrow$ individual model of subject j, with ϕ copied from θ
    Evaluate $\nabla_\phi \mathcal{L}(f_j)$ with respect to the same N shots and update ϕ of derived individual model f_j via $\phi - \alpha\nabla_\phi\mathcal{L}(f_j)$ ▷ Optional
  end for
end while

Meta-Testing

Infer (Z_x^k, Z_y^k) from N shots
$f_k(\cdot;\phi,\mathbf{Z}^j) \leftarrow$ individual model of subject k
Evaluate $\nabla_\phi \mathcal{L}(f_k)$ with respect to N shots and update with $\phi = \phi - \alpha\nabla_\phi\mathcal{L}(f_k)$ ▷ Optional

For each subject in both meta-training and meta-testing set, the framework takes few shots (queries) from it and infers its scaling factors (Z_x, Z_y), where x and y correspond to physiological and fiber responses' amplitudes respectively. Finally the subject's individual model, or individual fit in this case, is derived by rescaling the population fit accordingly. The framework uses Bayesian optimization (BO) [30]

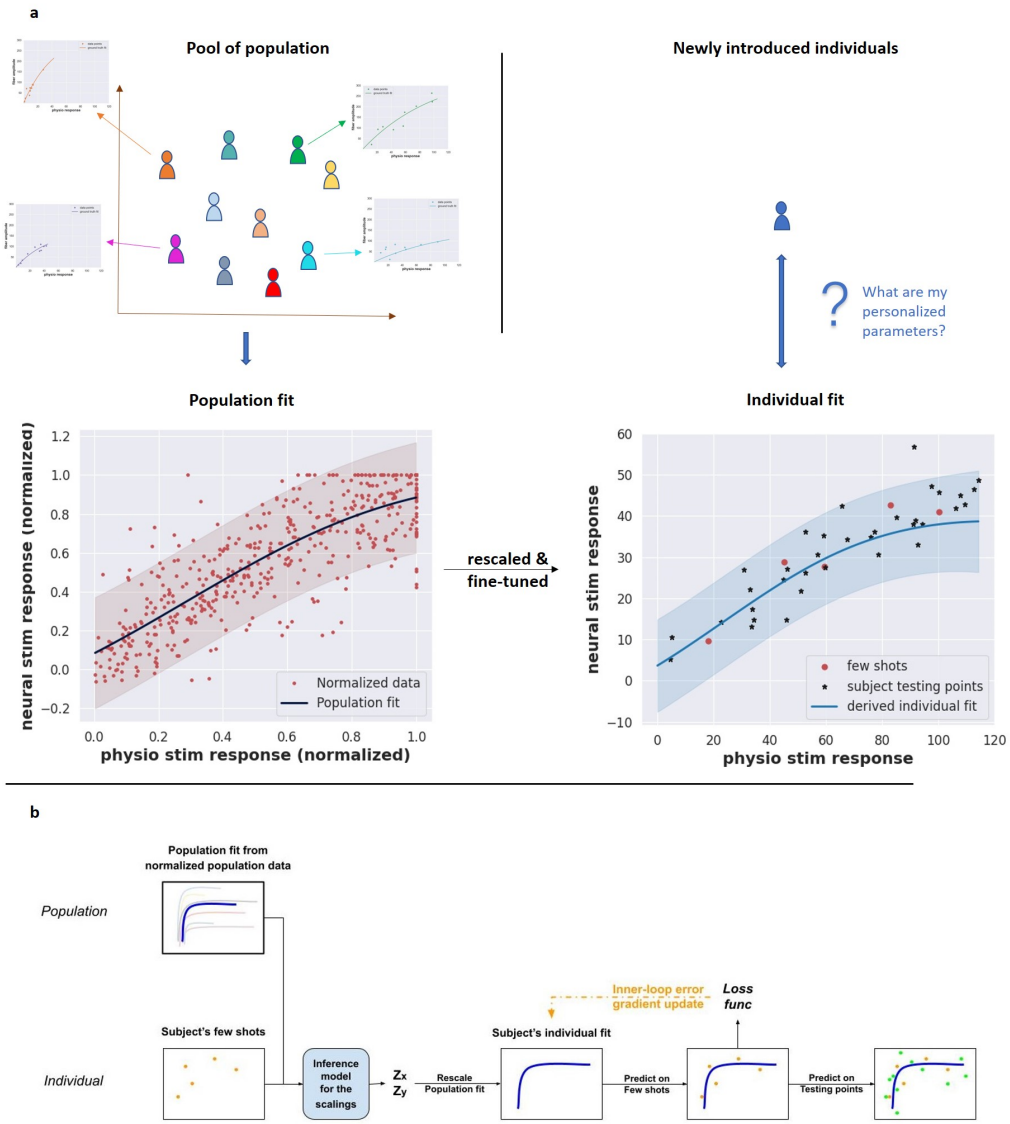

Figure 1: The proposed meta-learning framework to predict individual subject's fiber amplitudes. (a) Illustrations on utilizing the meta-learning framework to predict on newly introduced subject, where the objective is to learn a mapping between individual subjects' neural and physiological responses' amplitudes in response to stimulation. (b) Diagram of proposed meta-learning framework.

as the inference model, aiming to minimize the prediction error on the few shots by rescaled population fit, with respect to the scaling factors. We expect the resulted scaling factors to be proper estimates of the operating ranges of both physiological and fiber responses' amplitudes. Generalization performance of the individual model will be evaluated on the unseen data points of the same subject, which will be denoted as subject test points.

However, it imposes strong restraints to force population and individual models to differ by only the scaling. To cope with that, the framework incorporates the inner-loop error feedback, as an one-step gradient update based on the feedback on the few-shots. The gradient update is conducted on the parameters of each individual model, initially copied from the meta parameters of the population model. The inner-loop error feedback paths are illustrated in Figure 1(b) as yellow dashed lines. In addition, feedback on the population set's subject test points can also be used to improve the model's meta performance, which can be called as outer-loop error feedback. Here, for different evaluation setups that we are going to describe in the following section, we used the outer-loop error only to guide the search of the learning rate of the inner-loop gradient update, in a separate

training process beforehand, and we fixed the resulted learning rate during the actual evaluation. Since the outer-loop feedback path was not actively used during the evaluations, we exclude that here from the diagram in Figure 1(b).

Overall, the design principle of the framework is to transfer the prior knowledge embodied in the population model into the learning process, so that the fast adaptation of each individual model is feasible given only few data points. By choosing the shape of the population fit as the prior, we implicitly assume that each subject's individual fit is characterized mainly by its distinct scalings on both responses while keeping a similar base shape. Note that the assumptions are justifiable by the real data, for instance in [20], each fiber group shared the either linear or exponential shape across all the subjects. In the case when the obtained few shots of a subject do not agree with the population model, the introduction of the additional gradient update can aid in striking a profitable trade-off between placing confidence fully on the prior assumptions, which can be imperfect due to subject variations, and fully on the measurement data, which is likely noise injected.

2.1.1. Variants of algorithms under evaluations

By varying the choices of certain hyperparameters, our meta-learning framework defines a family of algorithms. In evaluating our approach, we consider four variants out of them: `metaRG` and `metaGP` have either regression or GP as the population models, where `metaRG-gu` and `metaGP-gu` would further enable inner-loop gradient update on top of their base versions.

2.2. Evaluation Metrics and Baseline Comparisons

Two evaluation metrics are used for performance comparison. The first is the test rooted mean squared error (RMSE), which is the prediction error of derived individual model on subject test points.

$$RMSE = \sqrt{E[(y - f(x))^2]}, \tag{1}$$

or the normalized version of it, relative RMSE (rRMSE), where the predictions were normalized by ground truth scaling factors.

$$rRMSE = \sqrt{E[((y - f(x))/Z_y)^2]}, \tag{2}$$

During evaluations, variants of the framework were compared with two baseline methods: `baseRG` and `baseGP`: closed-form regression and GP fitted directly on the subject few-shots. Baselines represent the case where knowledge of population is not explicitly utilized in deriving the individual fits via rescaling. Whenever available, we allowed the algorithms with regression population model (those with `RG`) use the best possible function forms. For example, the regression model will use a linear function if we know a priori the data demonstrates a linear relationship.

2.3. Performance Evaluations

We conducted three sets of evaluations. For the first set, we created from scratch purely synthetic subjects from three commonly-seen 1-dimensional functions, where we had access to all their ground-truth information. The purely synthetic datasets were built either "homogeneously", that is when all subjects shared the same function, or "heterogeneously", that is when more than one functions involved. With this set of evaluations, we can verify the performance of meta-learning in capturing this population knowledge.

The second set of evaluation was conducted on publicly available real experimental data, from previous publications [20], where we can observe the performance of meta-learning in practical settings.

However, evaluations on the real dataset were limited by its small size, preventing us to investigate the asymptotic performance with respect to the number of shots. In order to gain more insights in potential large-scale practical settings, we synthetically expanded one real dataset as the third set of evaluations. We termed this one as semi-synthetic, highlighting its connections with the real data. Other than being used in the comparison of algorithms, the semi-synthetic datasets was also used to investigate how different population sets affect few-shot prediction performance. Specifically, these datasets allow to probe the effect of population size, variability and number of shots per individual. In the following, we will describe in order the evaluation results on the three datasets.

2.3.1. Evaluations on Purely Synthetic Datasets

Figure 2 shows the evaluation results of all six algorithms on one purely synthetic dataset named `syn-exp`, a homogeneous dataset where all its subjects were generated from exponential function. Please refer to the Section 4.5 for details regarding subject generation and exponential function specifications. We performed leave-one-out cross-validations and reported the average testing rRMSE per subject in Figure 2(b) and (c) (where (c) is the zoom-in view of (b) when number of shots is from 5 to 20). These plots show, in general, variants of the framework all demonstrated competitive performance, with most of them outperformed their counterparts in the baselines, especially in the case when the number of shots was small. In addition, we included also the red dashed line as the dataset noise, a scalar calculated by averaging the ground truth noise standard deviation of all the synthetic subjects. Judging by the plots, all algorithms

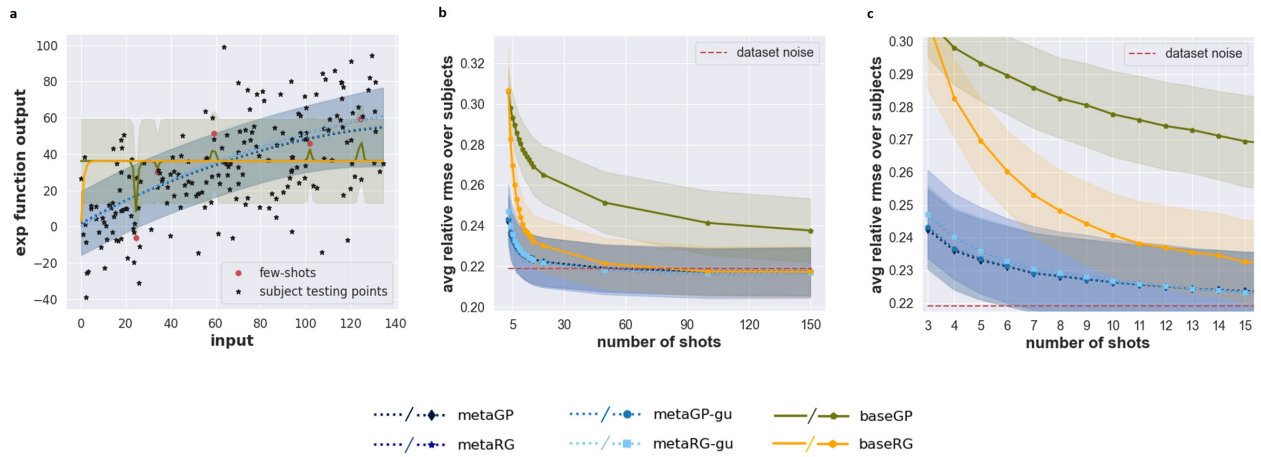

Figure 2: Performance comparison between the algorithms adopting the meta-learning framework and the baselines on a purely synthetic dataset built from exponential function `syn-exp`. (a) Examples of derived individual fits on 5-shots predictions. (b) Performance curve of relative rooted mean squared error (rRMSE) with respect to number of shots. (c) the same rRMSE curve as (b) but zooming in the region when number of shots is small.

approached it when more shots of the subject were made available. But the variants of the frameworks were always close to the dataset noise, even when there were only 3 to 5 shots available. This results showcase the benefits of utilizing the extracted prior knowledge from the meta-training phase, in relieving and even overcoming the lack of necessary information caused by only accessing to small size of data.

In addition, the rRMSE plots revealed a clear performance disadvantage in `baseGP` compared to the others, and it can also be seen clearly from Figure 2(a), where sample individual fits on 5-shot predictions are plotted, that GP alone did derive poor fits due to failure to recover the exponential form as well as overfitting to the few-shots. Although this show that the GP alone might not be a good choice for fitting few-shot individual model, combining GP with the meta-learning framework did lead to a definitive improvement regardless of the number of shots in the evaluations. This demonstrate that our proposed meta-learning framework is flexible in terms of the choice of the population model.

Aside from `syn-exp`, the evaluations were conducted on two other purely synthetic datasets: a homogeneous one of sinusoidal function and a heterogeneous dataset comprised of both exponential and linear subjects. Similar trends were observed across all three datasets, please refer to Appendix A for details of their results.

2.3.2. Evaluations on Vagus Nerve Fiber Recruitment datasets We used the vagus nerve fiber recruitment data from [20] as the real experimental datasets. The real data included data points describing three 1-dimensional physio-fiber relationships: from EMG to A-fiber, from ΔHR to B-fiber and from ΔBI to C-fiber, hereon denoted as `EMG-A`, `cHR-B`, `cBI-C` respectively. The same as the purely synthetic evaluations, leave-one-out cross-validations were adopted for performance comparisons on RMSE as no ground truth scaling factors are available.

The comparisons on the real experimental datasets are given in Table 2, Figure 3 and Figure 4. Figure 3 shows plots on the similar evaluations as the purely synthetic ones, with the plots in the left column include the sample individual fits from 3-shots predictions and those in the right column include the RMSE curve with respect to number of shots. Unfortunately, we were forced to limit the number of shots between 2 and 4, due to constraints on the dataset size. Nevertheless, we still observed general trends similar to the purely synthetic ones: performance improvements with more shots and the advantages of using the meta-learning framework over the baselines. It is evident that `baseGP` would fail the task of few-shot learning, and that the meta-learning remedied that making use of the prior knowledge, whereas `baseRG` performed similar yet in general worse to that of meta-learning variants. See Table 2 for a quantitative comparison on 3-shot prediction, listing the RMSEs of all the algorithms in all the datsets. Note that the best RMSE obtained are made bold in the table, and all of which were from variants of the framework. However, `baseRG` was clearly not an appropriate choice for 2-shots prediction in `cHR-B` and `cBI-C`, likely due to fitting near perfectly on only 2 data points, while none of the framework

variants displayed such severe degradation on the 2-shots. This is likely because the shape of population fit is providing a regularizing effect, preventing the individual fit to overfit the 2 shots. Last but not least, the competitive performance of combing GP with meta-learning is on par with those utilizing the regression forms carefully chosen for this dataset, meaning that GP becomes a viable choice in fitting the real data when using meta-learning.

Furthermore, we provide in Figure 4 the combined plots including all subjects' individual fits (averaged over 50 repetitions) for `metaGP-gu` and `metaRG-gu`. Judged by the plots, the meta-learning framework was able to capture the varieties in the subjects' operating ranges of both physiological and fiber responses' amplitudes, and their derived individual fits correlated well with the data points. As one additional point, because BO for `metaGP-gu` used mean negative loglikelihood, choosing the scaling factors would consider both mean and variance of the error. This could partially explain the conservative behaviours of not using a large scaling factor on the axis of fiber amplitudes, observed in some of its individual fits like subject 1 of `cHR-B`.

Table 2: Comparisons of algorithms' testing performance on all three regression tasks of the real experimental data. For each algorithm and each task, leave-one-out cross-validation was conducted, and the values reported were its average testing RMSE over all subjects. The best performance of each task is made bold in the table.

	`EMG-A`	`cHR-B`	`cBI-C`
metaGP	16.92	7.98	30.88
metaGP-gu	**16.66**	**7.82**	30.49
metaRG	16.999	8.08	**29.60**
metaRG-gu	17.002	8.02	29.64
baseGP	23.58	11.72	44.19
baseRG	17.03	9.41	35.71

2.3.3. Semi-Synthetic Datasets We created semi-synthetic datasets `semi-C-syn` as synthetic expansion of real dataset `cBI-C`. The procedure of generating semi-synthetic subject was similar to that of purely synthetic subject, except for that certain statistics were coming from the real data. Please refer to Section 4.5 for details. The results on semi-synthetic datasets are shown in Figure 5. In Figure 5(a) and (b), we can again observe the similar trends on the derived individual fits and on the algorithm performance with respect to the number of shots. There are two potential reasons that could account for the gap between the best performing algorithm in the plot and the dataset noise. One is the insufficient number of shots, and the other is the difference among the true generating functions for the synthetic subjects.

To conduct the investigations on how different population sets affects few-shot prediction performance, we further created multiple semi-synthetic datasets, to be utilized purely as the population set. The bottom part of Figure 5 demonstrates how varying aspects of the population set influenced the performance of the meta-testing set of `semi-C-syn`. Here we show the results of two investigations, and in each of which we concentrated on either the number or the noise level of the subjects in the population set, referred to as `ns` and `l_ni` respectively. We adopted the same color heat map, shown in side with the plots, to facilitate observing the trends of varying that specific hyperparameter. Judging from the results both its total size and its noise level are important factors to care in the population set. In practice, noise is unfortunately not directly manipulable, but this evaluation results indicate that increasing the size can be an effective countering measure. We conducted also the investigations on other hyperparameters, namely `np` the number of data points per subjects and `y_nm` range (normalization) of subjects' fiber amplitude, but neither was found to be as crucial as above two. Please see Appendix B for details of their results.

3. Discussion

The impacts of the meta-learning framework are two-fold. For the fiber amplitudes prediction problem itself, it serves as a pragmatic decision making tool to investigate the relationship between fiber and physiological responses, and it can perform in the domain when amount of data for that subject is scarce. The compatibility with the scarce data is of significance as even obtaining one data point can be a challenging and time-consuming process in the real experiments. Recall that to collect one data point, the procedure involves the setting of stimulation parameters, delivery of neurostimulation, recording concurrently of neural and physiological signals, as well as any effort in post-analysis. In addition, due to the safety concern for the subject, the ability of having less stimulation to achieve the same objective is generally preferred. Hence it is beneficial to keep the repetitions of that process to a minimum. Furthermore, the framework allows us to probe the trade-off between the time and the performance. As shown in the evaluation results, more shots will lead to an improvement on the prediction error. Therefore, the framework offers the flexibility to adjust emphasis according to needs and constraints.

In this concrete problem of VNS, the core idea of the meta-learning framework is to leverage the shape of

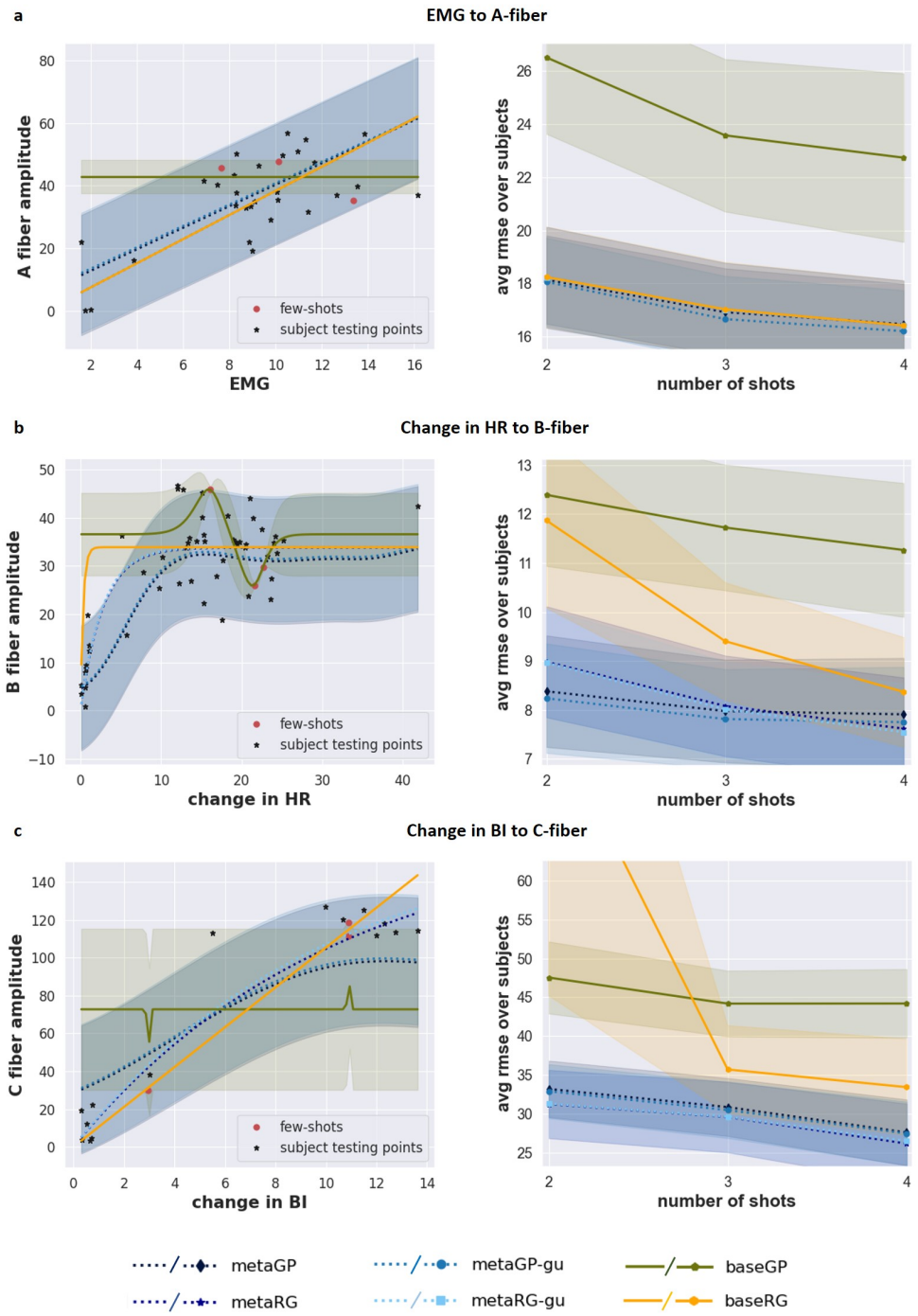

Figure 3: Performance of the meta-learning framework on vagus nerve stimulation fiber recruitment data. Each row includes examples of derived individual fits on 3-shots predictions (left) and rooted mean squared error (RMSE) comparison (right) with respect to number of shots. (a) Results on dataset `EMG-A`. (b) Results on dataset `cHR-B`. (c) Results on dataset `cBI-C`.

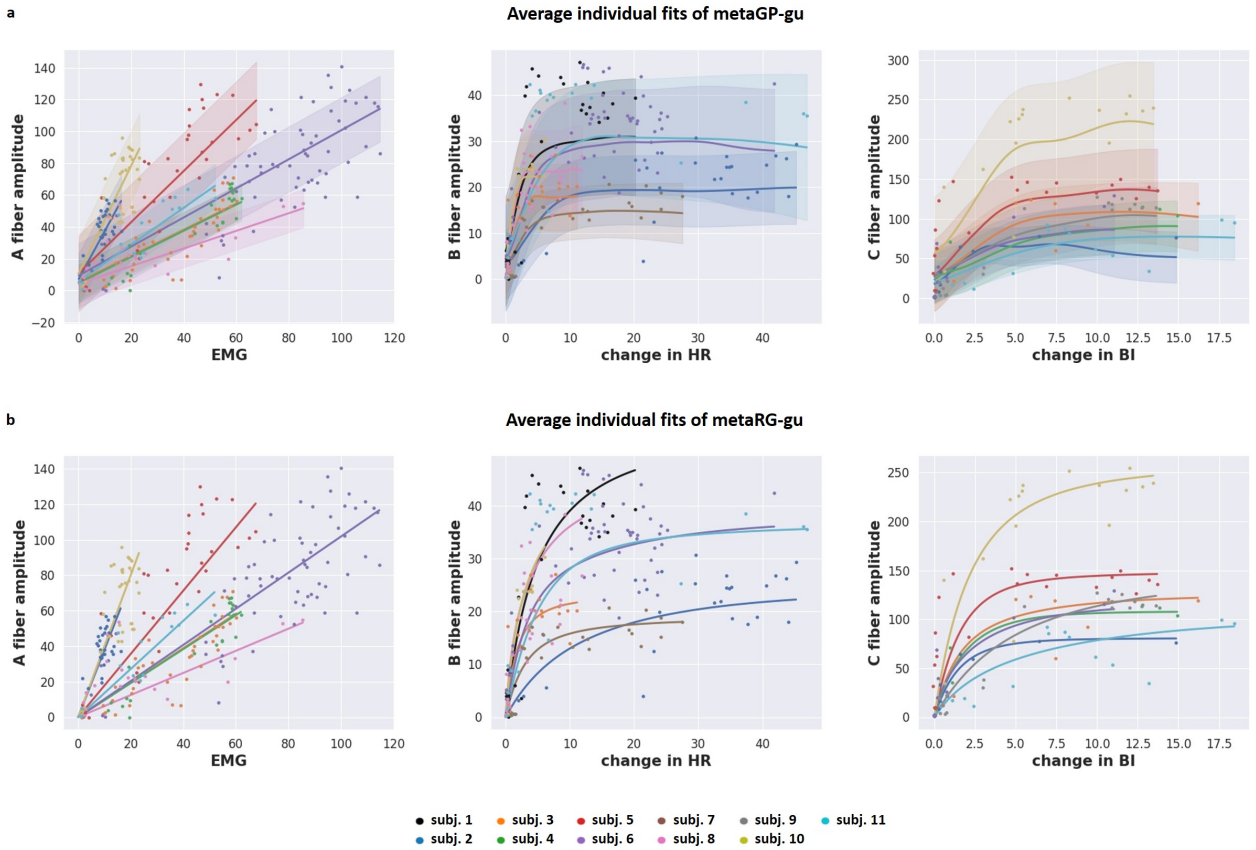

Figure 4: Derived individual fits on 3-shots predictions for all subjects in all three real datasets, from two variants of the meta-learning framework: (a) with GP as population model and with inner-loop gradient update (`metaGP-gu`); (b) with closed-form regression as population model and with inner-loop gradient update (`metaRG-gu`). Plotted are average mean over 50 repetitions, and average standard deviation on the GP prediction is shown for `metaGP-gu` only.

the population fit in the few-shot prediction of individual fiber recruitment by learning the subject-dependent operating ranges. This in general conveys a notion of knowledge transfer from the population to each new individual. Exploring such notion has benefits in a more broadly field of neurostimulation, since the challenges of personalization exists universally in different modalities. And our proposed meta-learning framework can be flexibly adapted to operate in various modalities, by re-defining prior knowledge to be extracted from the population and the key subject-dependent parameters according to the problems at hand. For example, the population model can be a convolutional model handling the image inputs or a recurrent model for time series signals, and the key individual parameters can be a specific weight matrix or length of time horizon.

The described meta-learning framework for VNS fiber recruitment prediction problem includes the option to utilize inner-loop feedback, in the form of one-step gradient update. In general, we can characterize the algorithms that use gradient update as the middle ground between placing full confidence in the prior knowledge from population and placing full confidence in the obtained few-shots from each individual. The former corresponds to `metaGP` and `metaRG`, while the latter is functionally closer to `baseGP` and `baseRG` with a good initialization if we allow sufficient gradient steps until convergence. By performing the gradient update properly, the derived individual model has the potential to not only attend to the distinctness of each subject but also avoid overfitting to the few shots. The latter is especially important for neural measurements, which commonly have high noise ratio. In our implementation of the framework, we adopted one-step gradient update. Although the number of steps was not carefully tuned, the evaluation results justified the choice. However, the number of gradient update should be viewed as a domain-sensitive hyperparameter, and the exact number is subject to change for various problems and modalities.

A key advantage of the proposed meta-learning

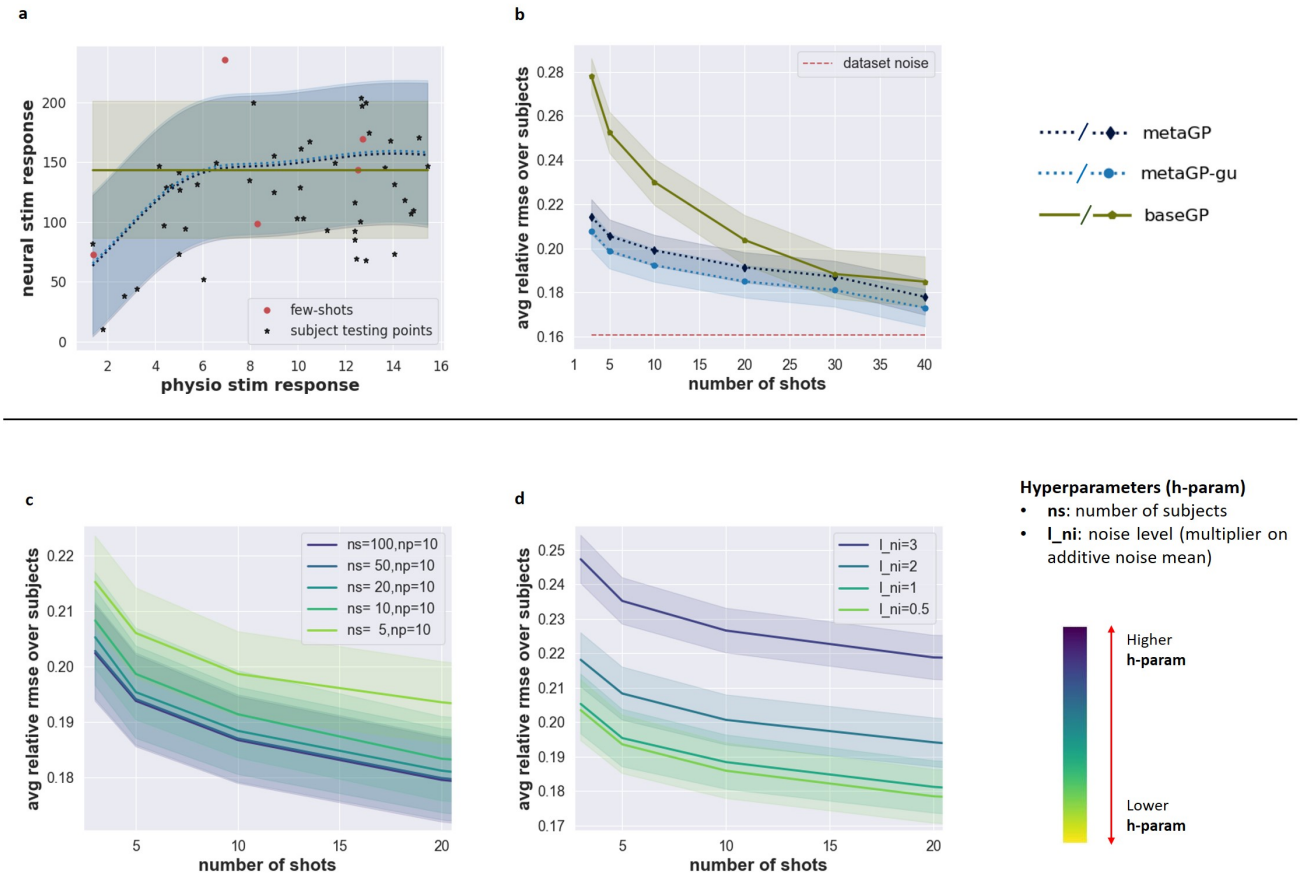

Figure 5: Performance comparison on semi-synthetic datasets `semi-C-syn`. (a)-(b) Examples of individual fits and testing relative rooted mean squared error (rRMSE) comparison for all algorithms with Gaussian process as population model. (c)-(d) Investigations on the effect of population set to testing rRMSE, with variations in two hyperparameters: (c) varying the number of subjects in the population set; (d) varying the noise ratio of the population set

framework is its compatibility with GP as a generic task-agnostic population model for any future data. Whenever a good closed-form regression model is easily obtainable, it is ideal to use it for better performance. For example in the case of dataset `syn-exp`, the exponential function is indeed the true data generating functions therefore the optimal choice for population model. But having access to the good regression model is not always viable, and even it is, the required time dedicated to search for it needs to be taken into consideration. In comparison, GP is advantageous in terms of flexibility, since the shape of GP varies with the provided data. GP also measures the uncertainty on the prediction, which has a good synergy when dealing with high noise ratio neural measurements. The choice of population model also goes beyond just GP and closed-form regression, in fact, any model with a tangible fit is directed applicable.

3.1. Limitation and Future Work

In its current form, our proposed meta-learning framework works the best for the task whose data is in general unimodal, since we assume that the shape of the learnt population model is the extracted prior knowledge from the task. This can impede the future applicability of the framework, when the task is highly bimodel or multimodal, for example a mix of cubic and sine functions, in which case learning a single population model could be not instructive anymore. Modifications of the framework to accommodate bimodel and multimodel in the population models are one future direction.

Up to now we have been focusing on the correlation across different subjects in one task, and another direction to pursue is to investigate on the correlation across different tasks for one subject. Correlation among different fiber groups of the subject was possible in the currently used real dataset, because the amplitudes of A-, B- and C-fibers were all measured

from eCAPs [20]. Although it remains unclear if learning on one fiber group would affect the others, multitask learning [24] can still be leveraged for such attempts.

The explicit designation of the form of prior knowledge (the shared shape) and subject-dependent knowledge (the scaling factors) can be replaced by autonomous discovery mechanisms. For example, we can formulate the inference as representational learning [23], that is to learn latent representations of a specific subject, and train the inference model jointly with the population model, as well as the individual model.

Beyond regression problems, we can investigate the use of the meta-learning framework on other problem types, such as the searching problem for optimal stimulation parameters using BO as well as the sequential decision problem for closed-loop systems using optimal control and reinforcement learning [25, 26].

4. Methods

4.1. Problem Formulation for Few-shot Regression

In a typical supervised learning [28] setup of machine learning, we learn a model on a training dataset D_{train} and evaluate its generalization performance on a testing dataset D_{test}. On the contrary, in meta-learning [9], we consider two datasets: meta-training set $\mathcal{D}_{meta_train}$ and meta-testing set $\mathcal{D}_{meta_test}$, each of which includes multiple regular datasets with its own D_{train} and D_{test} .

We can formulate the few-shot prediction for individual subject's fiber recruitment as k-shot regression, where the data points from the same subject construes a regular dataset $D \in \mathcal{D}$. For each subject, k physiology-fiber response pairs are made available to the model as D_{train}, while the rest are used for evaluation as D_{test}. For the typical few-shot learning setting, the value of k is small, for example $k <= 5$, while the larger value of k can be used to investigate the asymptotic performance. We consider the split of meta datasets by subject indices, meaning any subject is in either $\mathcal{D}_{meta_train}$ or $\mathcal{D}_{meta_test}$ but not both, this it to minimize the information leakage from meta-training to meta-testing.

4.2. Related Work on Personalized Treatment and Meta-Learning

4.2.1. Personalized Treatment Instead of providing the same treatment for all the patients, personalized medicine aims to make appropriate adjustments for each individual, and it is a widely encountered and investigated topic in neurostimulation [1, 36–38]. For example in [1], the treatment devices for spinal cord stimulation were reprogrammed to tailor to each patient's preference, and in [37], personalized information was incorporated into the BO algorithm as an extra input dimension for brain stimulation. The importance of personalized treatment is also highlighted in VNS, as the desirable stimulation parameters can be challenging to find due to subject variability in response to stimuli and various physiological factors [6, 17, 20, 22].

4.2.2. Meta-Learning In the recent years, meta-learning for few-shot learning was widely studied on applications including but not limiting to computer vision [9], natural language processing [39], robotics [40] and drug discovery [41]. One taxonomy approximately categorize meta-learning algorithms into three classes: metric-based, optimization-based and model-based [42]. In general, metric-based algorithms map the input to a common feature space and make the prediction in the sense of weighted average. The weights correlate with the metrics calculated between pairs of inputs, measuring the similarities between them. For example for Matching Network [43], softmax of the cosine similarity between the embeddings of two inputs is used as weight on the predictions. On the other hand, optimization-based algorithms aims to learn as similar as an optimization process, and the challenges are to answer how the gradient calculated on the few-shots can lead to a good solution in a few optimization steps. An example of optimization-based algorithm is MAML (Model-Agnostic Meta-Learning) [10], which tries to learn from meta-training set a good initialization on the model parameters, so that as small as one gradient update can let it adapt to a new task. Lastly for model-based algorithms, specially designed model are usually utilized for learning. In this category, MANN (Memory-Augmented Neural Network) [44] utilizes explicit external memory module to memorize the information and to facilitate fast learning of a new task. There is also natural connection between meta-learning and hierarchical Bayesian modeling, which utilizes a hierarchical form to infer the latent variables via Bayesian method [45], if viewing the meta-parameters and the model parameters as different level in the hierarchy. For example, [46] identified MAML can be seen as a point estimate version of empirical Bayes, a hierarchical Bayesian model whose parameters' prior distribution is estimated using the data, and in [47], the authors showed the derived PAC-bayes bounds for meta-learning was similar in the expression as the one for hierarchical variational Bayes.

Note that our proposed meta-learning framework is closer to the optimization-based algorithms, where we extracted the learnt population model as an "initial

point" for each individual model and incorporated gradient graduate from both inner-loop and outer-loop error feedback. To the best of the authors' knowledge, there hasn't been work investigating the usage of few-shot learning and meta-learning for VNS and neurostimulation applications.

4.3. Gaussian Process

GP can be thought as a generalization of Gaussian probability distribution that describes functions instead of scalars or vectors [27]. GP is a Bayesian method, and it defines a posterior distribution over functions given the prior as well as the data. The prior represents the beliefs we have over the kind of function we are looking for, before even seeing any data. The belief will be updated and refined after observing data points, which leads to the posterior distribution. For example, the posterior can put more preference on functions that better explains the observations. In addition, GP is a non-parametric model [8], whose capacity to fit increases along with more data points. When working with GP, we generally do not need to worry about underfitting the data, thus it is very flexible to use on any problem.

We consider GP for regression in our problem. Take a zero mean GP in the following form

$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}^{'})) \tag{3}$$

where $k()$ defines the kernel function, which is the covariance function between two inputs $\mathbf{x}$ and $\mathbf{x}^{'}$ representing the covariance between their outputs. Therefore, given multiple training inputs $\mathbf{X}$ and its observed outputs $\mathbf{y}$, assuming zero mean additive noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$, the predictive mean and covariance of testing inputs $\mathbf{X}^*$ can be written as

$$\mathbf{f}^*|\mathbf{X}^*, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{f}^*}, \mathbf{\Sigma}_{\mathbf{f}^*}) \tag{4}$$

$$\mu_{\mathbf{f}^*} = K(\mathbf{X}^*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1}\mathbf{y} \tag{5}$$

$$\begin{aligned}\mathbf{\Sigma}_{\mathbf{f}^*} =& K(\mathbf{X}^*, \mathbf{X}^*) - \\ & K(\mathbf{X}^*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} K(\mathbf{X}, \mathbf{X}^*)\end{aligned} \tag{6}$$

where $K()$ is the covariance matrix between two sets of points. Note that both Equation (5) and Equation (6) contains matrix inverse operation which has $O(N^3)$ complexity [29], where N is the number of entries in each dimension of matrix. As a result, GP is in general effective in small data domain but suffers from computational issue when a large data size is involved.

4.4. Bayesian Optimization

BO [30] is a method to find the maximum of a target cost function, especially practical when the one in question is expensive to evaluate. It does not rely on gradient, hence it is very useful when we do not have access to that information of the cost function. BO estimates the unknown cost function with a surrogate function, and like GP, it obtains posterior by updating the prior beliefs based on evidence, which is the responses of the function on multiple query points. Being both Bayesian methods, BO and GP share obvious synergy, and in fact, GP is a common choice of BO's surrogate function. BO is an iterative process. At each iteration, after deriving the new posterior, BO utilizes an acquisition function to determine where to query the target function at the next step. The role of an acquisition function is to balance exploration and exploitation aspects when searching for the next query point. The exploration is to query the high uncertainty region, where we haven't thoroughly investigated yet, and the exploitation is to query the high reward region, where we already observed higher values of responses from the function. By designing acquisition function to strike a profitable trade-off between exploration and exploitation, BO can keep the necessary queries to the function to a small number, which is essential when evaluating the target function is expensive.

The meta-learning framework applies BO to infer the scaling factors for each subject on both physiological and fiber responses' amplitudes, based on its given k shots. For our 1D regression problem, we designed a scalar cost function for each candidate pair of (Z_x, Z_y). The cost function evaluated a scaling factors pair, by measuring the prediction error of the derived individual fit, obtained by rescaling the population fit with the said pair. We can write the corresponding optimization problem as:

$$\max_{(Z_x, Z_y)} C(Z_x, Z_y) = E[-L(f(\mathbf{x}; Z_x, Z_y), \mathbf{y})] \tag{7}$$

where $f(\mathbf{x}; Z_x, Z_y) = Z_y f_{ppl}(\mathbf{x}/Z_x)$, $f_{ppl}()$ and $f()$ represent population model and derived individual model, and the cost function $L()$ can take the form of RMSE and mean negative log-likelihood.

We adopted an improvement-based acquisition function, Expected Improvement (EI) [30,31], to guide the search for the next query. EI works by finding the new query point by maximizing the expected improvement function

$$\max_{(Z_x, Z_y)} E[\max\{0, C_{t+1}(Z_x, Z_y) - C(Z_x^+, Z_y^+)\}] \tag{8}$$

where (Z_x^+, Z_y^+) is the best scaling factors pair found so far. Furthermore, EI in Equation (8) can be evaluated analytically assuming probability density function and cumulative density function of the standard normal distribution [31].

4.5. Synthetic Data Generation of Purely Synthetic and Semi-Synthetic datasets

Algorithm 2 describes the synthetic subject generation procedure for both purely synthetic and semi-synthetic subjects.

Algorithm 2 Data Generation Routine for Synthetic Subject i

Input:
generating function f
distribution for noise standard deviation $\mathcal{P}_\sigma$
distribution for scaling factor pair $\mathcal{P}_{(Z_x,Z_y)}$
number of data points n_s
Output: synthetic dataset D^i

initiate D^i as empty
sample $(Z_x^i, Z_y^i) \sim \mathcal{P}_{(Z_x,Z_y)}$
sample $\sigma^i \sim \mathcal{P}_\sigma$
for j=1,...,n_s **do**
  sample $x_j^i \sim \mathcal{U}_{[0,Z_x^i]}$
  sample $\epsilon_j \sim \mathcal{N}(0,(\sigma^i)^2)$
  $y_j^i \leftarrow Z_y^i f(x_j^i/Z_x^i) + \epsilon_j$
  add (x_j^i, y_j^i) to D^i
end for

4.5.1. Purely Synthetic Datasets Each subject in the purely synthetic datasets was generated by one of the following commonly-seen functions in Equations (9), (10) and (11), denoted as `exp`, `sin` and `lin` respectively.

$$y = A(1 - \exp(x/W)), \tag{9}$$

$$y = (A/2)\sin(x/W) + (A/2), \tag{10}$$

$$y = Ax, \tag{11}$$

where A and W are parameters of each function, corresponding to Z_y and Z_x in Algorithms 2 respectively. By varying values of A and W, the generation routine created synthetic subjects with different operating ranges on both input and output axis.

The `syn-exp` purely synthetic dataset contained 20 synthetic subjects all generated from `exp`, where each subject had 200 data points. Whenever few-shots were needed, they were sampled based on the amplitudes of input, meaning one sample for each non-overlapping fixed-length interval of the input domain. Note that we introduced this practice so that the sampled few-shots were more evenly distributed and not clustered. In addition, we also created another homogeneous dataset from `sin` with the same size as `syn-exp` and a heterogeneous dataset with 30 synthetic subjects, where half of them were generated by `exp` and the other half by `lin`. There were 100 data points per subject in the heterogeneous dataset. Please refers to the appendix for the evaluation results on the latter two datasets.

4.5.2. Semi-Synthetic Datasets As synthetic expansion of real dataset `cBI-C`, we used the statistics calculated from the real data to generate semi-synthetic subjects, with the aim that the generated semi-synthetic subjects are reasonable to appear in practice:

- The generating function for each subject f^i was sampled from population GP fitted in `cBI-C`.
- Scaling factor Z_x^i was sampled from Gaussian distribution, with μ_x^c and σ_x^c being the empirical mean and standard deviation of maximum physiological responses across real subjects. The same strategy was applied to Z_y^i on fiber amplitudes.
- Noise standard derivation for each subject i was computed as $\sigma^i = Z_y^i * r^i$ where the process to obtain r^i was described as follows. Firstly, regression with `exp` was fitted for each real subject m. In other words, we fitted for each subject an exponential function as in Equation (9) using all its data points. Secondly, also for each real subject m, we recorded the ratio of the training error to the maximal range, written as $RMSE^m/Z_y^m$, as an approximate noise-to-signal ratio. Finally, r^i was sampled from another Gaussian distribution, with μ_r and σ_r being empirical mean and standard deviation of the above-mentioned ratio across real subjects.

To bear resemblance to practical situations, we kept the number of data points per subject relatively small when creating the dataset, whilst including a larger number of subjects. Consequently, the first semi-synthetic dataset, `semi-C-syn`, contained 50 synthetic subjects and 50 data points per subject. The real subject might has even less number of points, but we wanted to investigate the asymptotic trends with respect to the number of shots. Sampling scheme of semi-synthetic datasets for few-shots was same to that of purely synthetic datasets, and 5-fold cross-validation was adopted when computing the evaluation metrics.

Additionally, we constructed a another set of semi-synthetic datasets, with variations in four hyperparameters: number of subjects (`ns`), number of data points per subjects (`np`), stimulation response noise level (`l_ni`) and range of fiber amplitude (`y_nm`). Among the four, the former two control the population set in quantity, while the latter two correspond more to the variations in quality. In particular, we created a base semi-synthetic dataset of 110 subjects, with 10 data points per each subjects, and for different `ns`, we sampled different numbers of semi-synthetic subjects, from 5 to 100. As for different `l_ni`, we recreated a similar base dataset as the `ns` case, but replacing noise standard deviation σ with $\lambda(Z_y * r)$, with λ as the additional multipliers ranging from 0.5 to 3, then we sampled 20 subjects from each noise level. The similar procedure was repeated for `np` and `l_nm`, where their base dataset had 160 subjects, each with 20 data points. For `np`, we sampled different combinations of

subject number and points number per subject, while maintaining the total data points fixed as 300. And for `y_nm`, we added different multipliers, from 0.5 to 5, on ranges of Y-axis (fiber amplitude axis).

4.6. Brief Overview of Real Experimental Data

The utilized real experimental data is a publicly available one introduced in [20]. The data was from rat experiments, and eCAPs were recorded while registering three acute physiological markers to the VNS. There were three 1-dimensional pairs of data `EMG-A` (relating EMG and A-fiber), `cHR-B` (relating changes in HR and B-fiber) and `cBI-C` (relating changes in BI and C-fiber), and the authors that paper identified a linear relationship for `EMG-A` and exponential ones for `cHR-B` and `cBI-C`. All three datasets were coming from 11 rat subjects, and the average data points per subject was 32.8 (ranging from 14 to 66) for `EMG-A`, 27 (9 to 54) for `cHR-B` and 12.8 (7 to 20) for `cBI-C`. Each of the datasets included data from 8 subjects. To keep the manuscript concise, here we only include the aspects of the real data pertaining to our evaluations, and for comprehensive descriptions regarding the data collection processes and further data analysis, please refer to the original paper.

Following the data preprocessing procedure in [20], we extracted same three 1-dimensional regression datasets. For the closed-form regression model, we used exponential relationship in the form of `exp` (Equation (9)) and linear of `lin` (Equation (11)). The sampled few-shots from each real subject had evenly spaced *charge-per-pulse*, which was a feature in the real dataset calculated by multiplying stimulus intensity by the pulse width. Since *charge-per-pulse* is a feature defined by the stimulation parameters, we can assume readily control on it in practice, and it is reasonable to sample the data point based on it.

4.7. Implementations

We implemented the proposed meta-learning framework in Python. PyTorch [32] packages BoTorch [33] and GpyTorch [34] were utilized to perform BO and GP respectively, whereas all kinds of closed-form regressions were optimized with `scipy.optimize.curve_fit` [35]. We adopted PyTorch's automatic differentiation engine `torch.autograd` for all gradient computations. All the evaluations were ran in CPU machine.

Acknowledgement

This work is supported by grants to GL from Canada CIFAR AI Chair program, the Canada Research Chair in Neural Computations and Interfacing, and the Fonds de Recherche du Québec (research scholar program), and a grant to SZ from United Therapeutics Corporation (MD, US). The authors would like to thank Theodoros Zanos, Numa Dancause, Marco Bonizzato, Tristen Deleu and Leo Choiniere for discussions on the project motivations and future directions, Lorenz Wernisch for suggestions on related field, and Eric Elmoznino and Alexandre Payeur for suggestions on manuscript structures.

Reference

[1] Brooker C *et al* 2021 ECAP-Controlled Closed-Loop Spinal Cord Stimulation Efficacy and Opioid Reduction Over 24-Months: Final Results of the Prospective, Multicenter, Open-Label Avalon Study *Pain Practice* **21** 680-691

[2] Okun M S 2012 Deep-Brain Stimulation for Parkinson's Disease *New England Journal of Medicine* **367** 1529-1538

[3] Ben-Menachem E 2002 Vagus-nerve stimulation for the treatment of epilepsy *The Lancet Neurology* **1** 477-482

[4] Edwards C A, Kouzani A, Lee K H and Ross R K 2017 Neurostimulation Devices for the Treatment of Neurologic Disorders *Mayo Clinic Proceedings* **92** 1427-1444

[5] Handforth A *et al* 1998 Vagus nerve stimulation therapy for partial-onset seizures: a randomized active-control trial *Neurology* **51** 48-55

[6] Ahmed U, Chang Y C, Zafeiropoulos S, Nassrallah Z, Miller L and Zanos S 2022 Strategies for precision vagus neuromodulation *Bioelectronic medicine* **8**

[7] Wang Y Q, Yao Q M, Kwok J T and Ni L M 2020 Generalizing from a Few Examples: A Survey on Few-shot Learning *ACM Comput. Surv.* **53** 1-34

[8] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press)

[9] Ravi S and Larochelle H 2017 Optimization as a Model for Few-Shot Learning *Int. Conf. on Learning Representations (ICLR)*

[10] Finn C, Abbeel P and Levine S 2017 Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks *Proc. Int. Conf. on Machine Learning (ICML), PMLR* **70** 1126-1135

[11] Sigrid B, Aleksandra K, Gerhard R and Gregor H 2018 Vagus Nerve as Modulator of the Brain–Gut Axis in Psychiatric and Inflammatory Disorders *Front. Psychiatry* **9**

[12] Rush A J *et al* 2005 Vagus nerve stimulation for treatment-resistant depression: a randomized, controlled acute phase trial *Biol Psychiatry* **58** 347-354

[13] George M S, Ward H E Jr, Ninan P T, Pollack M, Nahas Z, Anderson B, Kose S, Howland R H, Goodman W K and Ballenger JC 2008 A pilot study of vagus nerve stimulation (VNS) for treatment-resistant anxiety disorders *Brain Stimul.* **1** 112-121

[14] Merrill C A *et al* 2006 Vagus nerve stimulation in patients with Alzheimer's disease: additional follow-up results of a pilot study through 1 year. *J Clin Psychiatr* **67** 1171-1178

[15] Chakravarthy K, Chaudhry H, Williams K and Christo P J 2015 Review of the Uses of Vagal Nerve Stimulation in Chronic Pain Management *Current pain and headache reports* **19** 54

[16] Koopman F A *et al* 2016 Vagus nerve stimulation inhibits cytokine production and attenuates disease severity in rheumatoid arthritis *Proc Natl Acad Sci USA* **113** 8284-8289

[17] Labiner D M and Ahern G L 2007 Vagus nerve stimulation therapy in depression and epilepsy: therapeutic parameter settings *Acta Neurol Scand.* **115** 23-33

[18] Musselman E D, Pelot N A and Grill W M 2019 Empirically Based Guidelines for Selecting Vagus Nerve Stimulation

Parameters in Epilepsy and Heart Failure *Cold Spring Harb Perspect Med.* **9**

[19] He S M, Teagle H F B and Buchman C A 2017 The Electrically Evoked Compound Action Potential: From Laboratory to Clinic *Front. Neurosci.* **11**

[20] Chang Y C *et al* 2020 Quantitative estimation of nerve fiber engagement by vagus nerve stimulation using physiological markers *Brain Stimul.* **13** 1617–1630

[21] McAllen R M, Shafton A D, Bratton B O, Trevaks D and Furness J B 2018 Calibration of thresholds for functional engagement of vagal A, B and C fiber groups in vivo *Bioelectronics in medicine* **1** 21-27

[22] Qing K Y, Wasilczuk K M, Ward M P, Phillips E H, Vlachos P P, Goergen C J and Irazoqui P P 2018 B fibers are the best predictors of cardiac activity during Vagus nerve stimulation *Bioelectronic Medicine* **4** 5

[23] Bengio Y, Courville A and Vincent P 2013 Representation Learning: A Review and New Perspectives *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35** 1798-1828

[24] Zhang Y and Yang Q 2021 A Survey on Multi-Task Learning *IEEE Transactions on Knowledge and Data Engineering*

[25] Sutton R S and Barto A G 2018 *Reinforcement learning: An introduction* (MIT press)

[26] Bertsekas D P 2000 *Dynamic Programming and Optimal Control* (Athena Scientific)

[27] Rasmussen C E and Williams C K I 2006 *Gaussian Processes for Machine Learning* (MIT Press)

[28] Bishop C M 2007 *Pattern Recognition and Machine Learning* (Springer New York)

[29] Wilson A G, Hu Z, Salakhutdinov R and Xing EP 2016 Deep Kernel Learning. *Proc. Int. Conf. on Machine Learning (ICML), PMLR* **51** 370-378

[30] Brochu E, Cora V and Freitas N E 2010 A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning ArXiv:abs/1012.2599

[31] Mockus J, Tiesis V and Zilinskas A 1978 The Application of Bayesian Methods for Seeking the Extremum *Towards Global Optimization* **2** 117-129

[32] Paszke A *et al* PyTorch: An Imperative Style, High-Performance Deep Learning Library 2019 *Advances in Neural Information Processing Systems (NeurIPS)* **32** 8024-8035

[33] Balandat M, Karrer B, Jiang D R, Daulton S, Letham B, Wilson A G and Bakshy E 2020 BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization *Advances in Neural Information Processing Systems (NeurIPS)* **33**

[34] Gardner J R, Pleiss G, Bindel D. Weinberger K Q and Wilson A G 2018 GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration *Advances in Neural Information Processing Systems (NeurIPS)* **31**

[35] Virtanen P *et al* 2020 SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python *Nature Methods* **17** 261-272

[36] Kallewaard J W *et al* 2021 Real-World Outcomes Using a Spinal Cord Stimulation Device Capable of Combination Therapy for Chronic Pain: A European, Multicenter Experience *J. Clin. Med.* **10**

[37] van Bueren N E R, Reed T L, Nguyen V, Sheffield J G, van der Ven S H G, Osborne M A, Kroesbergen E H and Kadosh R C 2021 Personalized brain stimulation for effective neurointervention across participants *PLOS Computational Biology* **17** 1-24

[38] Hollunder B, Rajamani N, Siddiqi S H, Finke C, Kühn A A, Mayberg H S, Fox M D, Neudorfer C and Horn A 2022 Toward personalized medicine in connectomic deep brain stimulation. *Progress in Neurobiology* **210**

[39] Brown T B *et al* 2020 Language Models are Few-Shot Learners *Advances in Neural Information Processing Systems (NeurIPS)* **33**

[40] Finn C, Yu T, Zhang T, Abbeel P and Levine S 2017 One-Shot Visual Imitation Learning via Meta-Learning *Proceedings of the 1st Annual Conference on Robot Learning, PMLR* **78** 357-368

[41] Altae-Tran H, Ramsundar B, Pappu A S and Pande V 2017 Low Data Drug Discovery with One-Shot Learning *ACS Central Science* **3** 283-293

[42] Hospedales T, Antoniou A, Micaelli P and Storkey A 2021 Meta-Learning in Neural Networks: A Survey *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1**

[43] Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, and Wierstra D 2016 Matching networks for one shot learning *Advances in Neural Information Processing System (NeurIPS)* **29**

[44] Santoro A, Bartunov S, Botvinick M, Wierstra D and Lillicrap T 2016 Meta-learning with memory-augmented neural networks *Proc. Int. Conf. on Machine Learning (ICML), PMLR* **48** 1842–1850

[45] Blei D, Ng A, and Jordan M 2003 Latent dirichlet allocation. *Jounral of Machine Learning Research* **3** 993–1022

[46] Grant E, Finn C, Levine S, Darrell T. and Griffiths T L 2018 Recasting Gradient-Based Meta-Learning as Hierarchical Bayes *Int. Conf. on Learning Representations (ICLR)*

[47] Amit R and Meir R 2018 Meta-Learning by Adjusting Priors Based on Extended PAC-Bayes Theory *Proc. Int. Conf. on Machine Learning (ICML), PMLR*

Appendix A. Additional Evaluations on Purely Synthetic Datasets

Here we show complete evaluations on purely synthetic datasets in Figure A1. For `syn-sin`, fitting regression model needed good initial guesses to be provided to the optimizer, as a result, all regression based algorithms performed very similarly. On the other hand, getting GP fit was much simpler, and with inner-loop gradient update, `metaGP-gu` could reach the desirable rRMSE, in the close proximity of that of regression based ones. For `syn-mix`, the heterogeneous nature of the dataset brought difficulties onto the learning, but not by a large margin, as both noisy data from `exp` and `lin` can have similar shape of the fits to some extends. From the results, despite a larger rRMSE gap from the dataset noise, the framework still outperformed baseline by a considerable margin.

Appendix B. Complete investigations on the effects of population set

Figure B1 demonstrates the complete investigations on how different population set hyperparameters influenced the performance in `semi-C-syn`, including the investigations on all four hyperparameters. The same color heat map as in Figure 5 was adopted to indicate the values of the hyperparameters. As mentioned above, judging from the results, the more important factors on the population set were its total size and its noise level.

The effect of `np` had a more complicated nature, as it affected not only the size of data from each subject but indirectly also the normalized population set. Recall that to obtain the population models, the data had to be normalized by its respective maximal values. During subject generation, the semi-synthetic datasets were injected with zero-mean Gaussian noise, so the absolute maximal fiber responses can be larger than the ground truth Z_y, meaning the normalization factor on fiber amplitude was commonly overestimated if `np` equals to the largest possible value. However, it is undesirable also to keep `np` too small, for example, if `np` is 1, every point in the normalized population set becomes $(1, 1)$. We would hypotheses an intermediate value as an optimal one, and Figure B1(b) suggested that it might be justified, but still uncertain due to the small gaps among the versions in comparison. Therefore, in addition, we computed firstly in Table B1 the mean and standard deviation of the RMSE distance between the fit used to generate all the subjects and the obtained population fit for various combinations of `ns` and `np`, and the results corroborate that the population set with an intermediate value of `np` did have a closer shape to the ground truth. Secondly, we ran the same evaluations of varying `np`, but this time with datasets of `y_nm` as 2, 3 and 5. Note that original one had `y_nm` as 1. The results are shown in Figure B2, and they show the similar trends as Figure B1(b) that intermediate value performs the best.

Lastly, it can be summarized that the least influencing factor was the range of fiber amplitude ranges, as varying that hyperparameter yielded almost identical curves. One plausible explanation is that since the population set was only utilized in a normalized fashion, the effect introduced by varying a multiplier on the amplitude ranges could be suppressed afterwards. This also demonstrate that the framework is resilient to a broad range of subjects varying the operating ranges of fiber amplitudes, which is a desirable feature in practice.

Table B1: Comparisons among different combinations of the number of subjects and the number of points per subject in the population set. Values reported are the distance between the data generating function and each obtained population fits

population set size	mean	std
`np`=150; `ns`= 2	0.066	0.023
`np`=100; `ns`= 3	0.055	0.013
`np`= 60; `ns`= 5	0.06	0.012
`np`= 30; `ns`=10	0.08	0.011
`np`= 15; `ns`=20	0.095	0.013

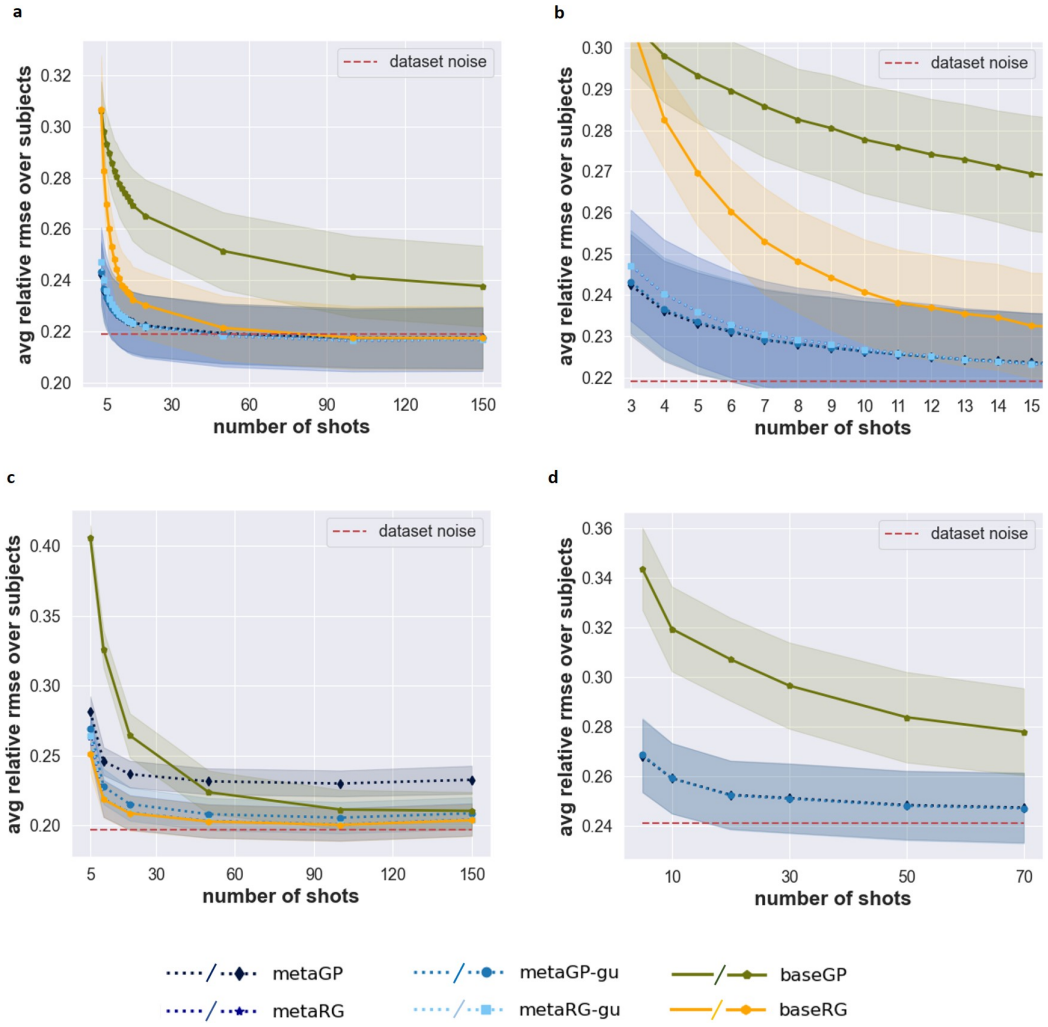

Figure A1: Complete performance comparison between the algorithms adopting the meta-learning framework and the baseline on three purely synthetic datasets. (a) Performance curve of relative rooted mean squared error (rRMSE) with respect to number of shots from exponential functions `syn-exp`. (b) Zoom-in of rRMSE curve of (a) when number of shots is small. Note that (a) and (b) are the same figures as Figure 2. (c) Performance curve of rRMSE with respect to number of shots on dataset built from sinusoidal function `syn-sin`. (d) Performance curve of rRMSE with respect to number of shots on dataset built from the mixtures of linear and exponential function `syn-mix`.

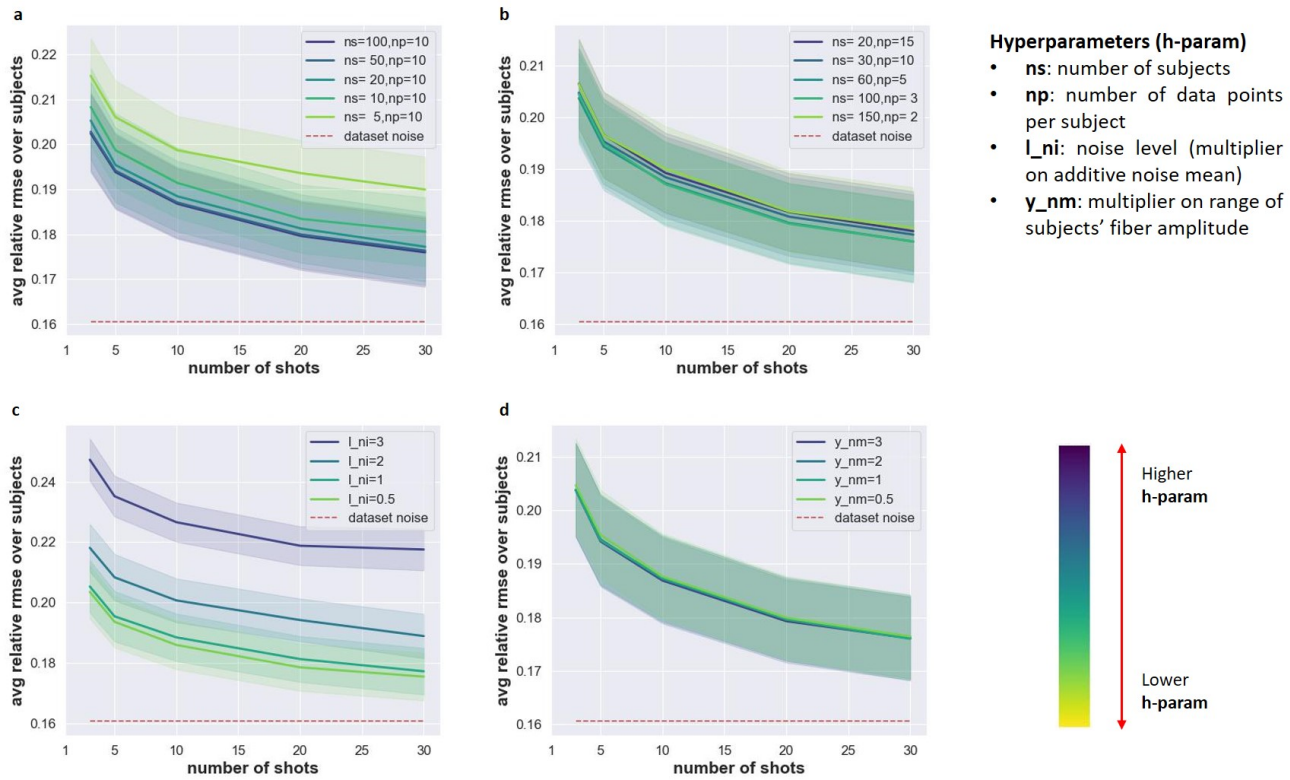

Figure B1: Investigations on the effect of the population set to testing relative relative rooted mean squared error, with variations in four hyperparameters: (a) varying the number of subjects in the population set; (b) varying the number of data points in each each subject while keeping the total size of population set fixed; (c) varying the noise ratio of the population set; (d) varying the operating range of the population set.

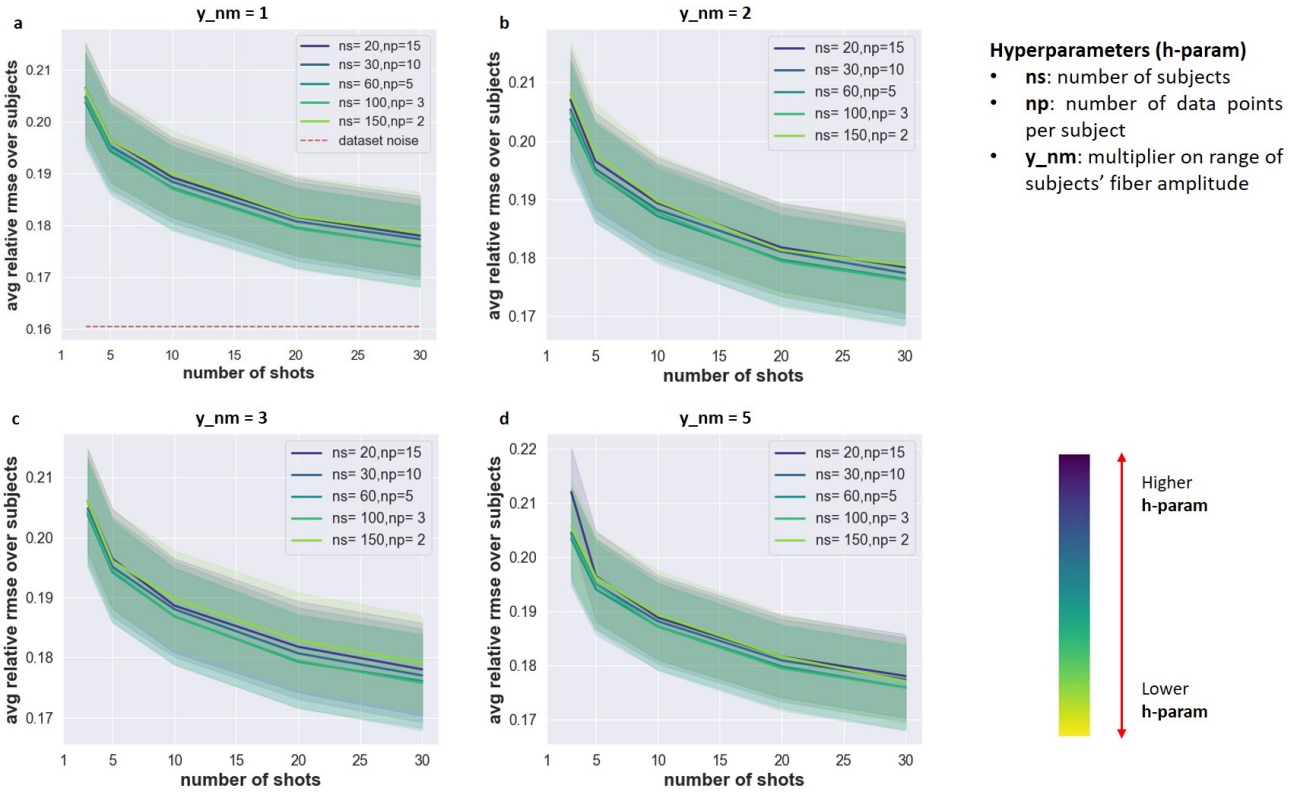

Figure B2: Investigations on the effect of `np` (number of data points per subject) with varying `y_nm` (multiplier on ranges of subjects' fiber amplitude of the population set). Each plots show relative rooted mean squared error: (a) `y_nm` as 1, this is exactly the same figure as in Figure B1(b); (b) `y_nm` as 2; (c) `y_nm` as 3; (d) `y_nm` as 5.