Supermodeling - a meta-procedure for data assimilation and parameters estimation

Leszek Siwik[1], Marcin Łoś[1], and Witold Dzwinel[1]

Department of Computer Science, AGH University of Science and Technology, Krakow, Poland

Abstract. The supermodel synchronizes several imperfect instances of a baseline model - e.g., variously parametrized models of a complex system - into a single simulation engine with superior prediction accuracy. In this paper, we present convincing pieces of evidence in support of the hypothesis that supermodeling can be also used as a meta-procedure for fast data assimilation (DA). Thanks ago, the computational time of parameters' estimation in multi-parameter models can be radically shortened. To this end, we compare various supermodeling approaches which employ: (1) three various training schemes, i.e., *nudging*, *weighting* and *assimilation*, (2) three classical data assimilation algorithms, i.e., ABC-SMC, 3DVAR, simplex method, and (3) various coupling schemes between dynamical variables of the ensembled models. We have performed extensive tests on a model of diversified cancer dynamics in the case of tumor growth, recurrence, and remission. We demonstrated that in all the configurations the supermodels are radically more efficient than single models trained by using classical DA schemes. We showed that the tightly coupled supermodel, trained by using the *nudging* scheme synchronizes the best, producing the efficient and the most accurate prognoses about cancer dynamics. Similarly, in the context of the application of supermodeling as the meta-algorithm for data assimilation, the classical 3DVAR algorithm appeared to be the most efficient baseline DA scheme for both the supermodel training and pre-training of the sub-models.

Keywords: supermodeling · data assimilation · tumor dynamics

1 Introduction – the concept of supermodeling

The assimilation of the computer model with a real phenomenon through a set of observations is a complex task and its time complexity increases exponentially with the number of parameters. This makes data assimilation (DA) procedures useless when applied for multiscale models such as models of weather dynamics [26,28] or biological processes like tumor growth [2]. Usually, such the multiscale models are highly parametrized.

It is well known from the literature that a multi-model ensemble produces more accurate prognoses than a single-model forecast [25]. Consequently, simultaneous estimation of the model parameters for the ensemble Kalman filter

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

(EnKF) is more efficient and accurate than for single data assimilation (DA) algorithm [3]. On the other hand, averaging trajectories from multiple models without synchronization may lead to undesired smoothing and variance reduction [7]. The alternative approach for taking advantage from the many trajectories followed by distinctive models and discovering many "basins of attraction" but without (premature) loss of the trajectories variety is combining models dynamically. The naive approach of this kind was proposed in [10]. The more mature solution was presented in [28] and consists in *dynamic combination of sub-models by introducing connection terms into the model equations that "nudge" the state of one model to the state of every other model in the ensemble, effectively forming a new dynamical system with the values of the connection coefficients as additional parameters*. In our previous paper [19] we posed a hypothesis

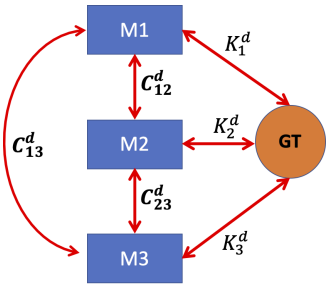


Fig. 1: The supermodeling coupling for 3 sub-models and *ground truth*.

that by combining a few imperfect, previously pre-trained, dynamical models and synchronizing them through the most sensitive dynamical variables one can overcome the problem of exponential inflation of the parameter space with their number [20]. Instead of estimating multiple parameters, one can train only a few coupling factors (hidden layer of data assimilation) of the model ensemble - the supermodel. In [19] we demonstrated that this meta-procedure can be more efficient than a single ABC-SMC classical data assimilation algorithm. However, in [19], we presented a case study, which uses only: (1) one supermodeling coupling scheme and (2) one classical DA algorithm. To make our hypothesis more credible we present here more extensive experiments.

The main contribution of this paper is the comparison of many versions of supermodels employing various training schemes and various classical DA algorithms as the baseline (i.e., for pre-training of the sub-models and training their coupling factors). Additionally, compared to our previous study, we employ a supermodeling scheme for a different and more realistic dynamical model, i.e., the model of tumor dynamics. The results of experiments allow for strengthening considerably the hypothesis about the benefits of using supermodeling as a meta-procedure for data assimilation for improving prediction accuracy.

2 Supermodel training schemes

Let us first define the supermodel assumptions [26]. Assuming, that we have $\mu = 1 \dots M$ sub-models, each described by the $d = 1 \dots D$ set of dynamical

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

variables, i.e.,

$$\dot{x}^d_\mu = f^d_\mu(x_\mu) \tag{1}$$

the supermodel set of equations looks as follows: [28]:

$$\dot{x}^d_\mu = f^d_\mu(x_\mu) + \underbrace{\sum^{\upsilon\neq\mu} C^d_{\mu\upsilon}(x^d_\upsilon - x^d_\mu)}_{\substack{\text{Synchronization}\\ \text{between sub-models}}} + \underbrace{K^d_\mu(x^d_{GT} - x^d_\mu)}_{\substack{\text{Synchronization}\\ \text{sub-model and the } \textit{ground truth}}} \tag{2}$$

where C_{ij} are the coupling (synchronization) factors between sub-models and K^d are the *nudging* coefficients attracting the models to the *ground truth* (GT) data (see Figure 1). We have assumed that $K^d = K$. The supermodel behavior is calculated as the average of the sub-models:

$$x^d_s(C,t) \equiv \frac{1}{M}\sum_\mu x^d_\mu(C,t). \tag{3}$$

Additionally, we define the synchronization quality as follows:

$$e^d(t) = \frac{1}{lp}\sum_{(\mu,\upsilon)}\frac{1}{N}\sum_{n=0}^{N-1}[x^d_\mu(n\Delta t) - x^d_\upsilon(\Delta t)]^2, \tag{4}$$

where N is the number of samples that discretize the trajectory, lp is the number of couplings between the sub-models, and Δt is the discretization interval. As shown in [27], the C tensor coefficients can be trained by using *ground truth* vector x_{GT} by minimizing the weighted squared error $E(C)$ in N following time-steps Δt; i.e.,

$$E(C) = \frac{1}{N\Delta t}\sum_{n=1}^{N}\int_{t_n}^{t_n+\Delta t}|x_s(C,t) - x_{GT}(t)|^2\gamma^t dt. \tag{5}$$

Error function $E(C)$ measures an accumulated numerical error, which includes the imperfections in the definition of the initial conditions. Discount value γ is from the (0,1) interval [27]. This decreases the contributing factors of the increases in the internal errors.

To develop a fully functional supermodel we have to (1) devise training algorithms for meta-parameters (C^d_{ij}, K^d_i), (2) select a proper set of the sub-models, (3) decide about the connections between them. In this section, we start with the training algorithms.

There exist a few different training algorithms for developing efficient and accurate supermodels. In the forthcoming sections we briefly discuss nudged, assimilated and weighted training schemes described in the previously published papers [28], [26], [19]. Many interesting novel concepts of the supermodels training, in the context of climate/weather forecast, are collected in [17].

2.1 Nudging supermodel

In this training scheme, during the first stage, the values of C^d_{ij} factors are updated alongside the sub-models according to the following formula [28]:

$$\dot{C}^d_{\mu\upsilon} = \alpha(x^d_\mu - x^d_\upsilon)(x^d_{GT} - x^d_\mu), \tag{6}$$

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

where α is a training constant. The second stage consists in running the coupled sub-models once again but this time C_{ij} are fixed to the values resulting from the first stage while the supermodel trajectory is *nudged* to GT data by correcting the values of K_i^d. The details of this procedure can be found in [27]. As the output of the *nudging* supermodel, we take the average of outputs of the M coupled sub-models obtained during the second stage of training.

2.2 Assimilated supermodel

This alternative approach, (which is referred to as *assimilated* supermodel) consists in pre-training of the sub-models and estimating coupling factors using well known data assimilation algorithms such as: Kalman filters [3], 3DVAR [23], Blue or EnsembleBlue [1], ABC-SMC [24] or optimization algorithms such as: Evolutionary Algorithms, Differential Evolution [15], Tabu Search [5,6] or classical derivative free optimization techniques [13,14] such as Simplex method [9]. Proposed idea consists in the following steps ([19]): (1) Apply a classical data-assimilation algorithm for a short pre-training of a few instances of the baseline model. (2) Create a supermodel from these imperfect sub-models coupled by only the most sensitive dynamical variables. Thus the number of coupling factors k will be small compared to the number of model parameters (e.g., for three sub-models coupled by only one variable $k = 3$). (3) Train these coupling factors using *ground truth* data by applying classical DA algorithm. The nudging coefficients are not required then. So now, the supermodel is described by the following equation:

$$\underbrace{\dot{x}_\mu^i = f_\mu^i(x_\mu)}_{\text{Pretrained submodels}} + \underbrace{\sum^{\upsilon \neq \mu} C_{\mu\upsilon}^i (x_\mu^i - x_\upsilon^i)}_{\text{Submodel coupling}} + \underbrace{K^i(x_{GT} - x_\mu^i)}_{\text{Synchronization submodels and ground-truth}} \tag{7}$$

What is important, in this procedure the most complex and time-consuming part, i.e., matching all the parameters of the baseline models, is significantly reduced due to a short pre-training only. Moreover, the instances (sub-models) can be created during only one pre-training, selecting the parameters corresponding to the best and the most diverse local minimums of the cost function. The sub-models can be also pre-trained parallelly reducing the computational time. In our experiments, shown in the Results section we have assumed that they are calculated in parallel.

The main part of the method is focused then on estimating (only a few) coupling factors. Since now, *nudging* to the *ground truth* is performed by the baseline data assimilation algorithm, so the part of equation 2, responsible for nudging the model towards the *ground truth* data, has been eliminated (see formula 7).

2.3 Weighted supermodel

The weighted sum of sub-models is the next and the most popular model ensembling approach. This time, the procedure is as follows: (1) Apply a classical

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

data assimilation algorithm for short pre-training of a few instances of the over-parametrized baseline models. (2) Ensemble the supermodel as the weighted sum of tentatively pre-trained imperfect sub-models with the weights w_μ matched to GT data by some DA algorithm. Thus the supermodel value x_s^d of the d_{th} dynamic variable is calculated now as:

$$x_s^d = \sum_{\mu=1}^{M} w_\mu x_\mu^d \tag{8}$$

The sum of weights w_μ is normalized to 1. As shown in [28] and [26] the weighted average scheme is equivalent to the *nudging* training scheme for large coupling factors.

3 Model of tumor dynamics

For the test case studies we use the model of the tumor (glioma brain cancer) dynamics: its growth and post anti-cancer treatment phases: remission and recurrence [16]. Unlike the *Handy* model considered in [19], the cancer model is more realistic and supported by real data. As shown by Ribba et. al [16] the model can be used in predictive oncology.

Proliferative cancer cells (P) represent fully functional cells that have the ability to multiply. It is also assumed that those cancer cells that are in unfavorable conditions (such as hypoxia or high mechanical pressure) transform into the quiescent cells (Q). Additionally, all of the proliferative and most of the quiescent cells die due to the anti-cancer drug activity (C). The surviving Q cells transform into mutated quiescent cancer cells (QP), which in turn convert to the proliferative (P) cancer cells (in reality even more aggressive ones). This model, though highly simplified, is a good metaphor for the principal processes influencing cancer dynamics during and after treatment. The model is described by the following set of ODEs [16]:

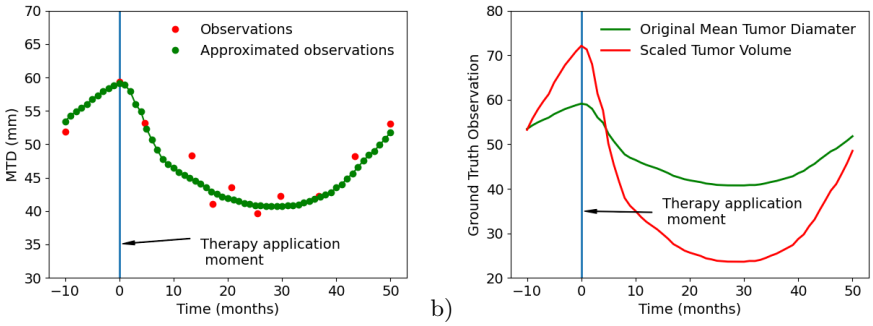
$$\begin{aligned}
\frac{\mathrm{d}C}{\mathrm{d}t} &= -KDE \times C \\
\frac{\mathrm{d}P}{\mathrm{d}t} &= \lambda_P \times P(1 - \frac{P^*}{K}) + k_{Q_pP} \times Q_P - k_{PQ} \times P - \gamma_P \times C \times KDE \times P \\
\frac{\mathrm{d}Q}{\mathrm{d}t} &= k_{PQ} \times P - \gamma_Q \times C \times KDE \times Q \\
\frac{\mathrm{d}Q_P}{\mathrm{d}t} &= \gamma_Q \times C \times KDE \times Q - k_{Q_PP} \times Q_P - \delta_{QP} \times Q_P \\
P^* &= P + Q + Q_P.
\end{aligned} \tag{9}$$

At a given moment of time t, the model state vector is $\boldsymbol{V}(t) = [P(t), Q(t), Qp(t), C(t)]$, where $P(t)$, $Q(t)$, $Qp(t)$ are the numbers of proliferating, quiescent and mutated quiescent cancer cells respectively. Assuming that the cancer cells have similar size, based on known average tumor density one can assume that the number of cells is proportional to the volume they occupy. Though, this will be rather a

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

rough estimation knowing the high variability of tumor density from the necrotic center to the tumor surface. The $C(t)$ dynamical variable represents anti-cancer drug concentration during chemotherapy or radiation dose in the case of radiotherapy. The approximate size of the tumor - calculated in the linear scale as MTD (the mean tumor diameter) - can be estimated from the total number of cancer cells $P^* = P + Q + Q_P$, and vice versa.

In [16], seven parameters were defined: λ_P – cell multiplication P; k_{PQ} – transition of P cells to into Q cells; γ_P – dying cells P; γ_Q – dying cells Q; KDE – therapy intensity; $k_{Q_P P}$ – transition of cells Q_P into P; and δ_{QP} – dying cells Q_P. For the initialization of the model, the initial values: P_0, Q_0, and Q_{P0} are given in [16].



a) b)

Fig. 2: Data augmentation via approximation of tumor dynamics for a patient with ID = 2 (see [16]) (a), and the comparison of the tumor dynamics measured by MTD and scaled tumor volume (b).

3.1 *Ground truth* data

For the experiments presented and discussed in this paper, the observations of the tumor dynamics for a patient, described in [16] as the patient with ID=2, have been adopted as the *ground truth* (GT) data. Our choice of the patient was completely random from almost 60 cases described in [16]. As shown in Figure2a, data represent the Mean Tumor Diameter (MTD) measurements in time. The Y-axis is scaled in millimeters (MTD) while X-axis (time) in months. The 0 value on X-axis means the start of anticancer therapy. Because glioma tumor has rather a regular shape, MTD can be relatively easy to measure and it reflects well the size of the tumor. In the case of irregular tumors, the cancer volume should be estimated instead. Because, the set of equations 9 describes the tumor time evolution in terms of the number of tumor cells, and assuming that this number is proportional to the volume V, one can easily estimate these volumes from MTD measurements. Therefore, MTD values taken from the Figure2a were converted into the volume of the tumor. Assuming that the tumor is a sphere (glioma is rather a cancer of regular spheroidal shape), the conversion has been made according to the equation: 10.

$$V = \frac{\pi (MTD)^3}{6} \tag{10}$$

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

The values of V obtained from 10 were scaled by dividing them by 1500 (arbitrary value) so that the volume and corresponding MTD values be of the same order of magnitude. The comparison of the tumor dynamics for the patient with the ID=2, measured in MTD, and corresponding tumor volumes are shown in Figure 2b. It is worth mentioning here that the authors of [16] made a mistake assimilating the equations directly to the linear MTD scale. That is why the final results from the paper [16] cannot be credible.

As shown in Figure 2a, for the purposes of this paper we made a simple data augmentation through the approximation of the measured data (60 points in green in Figure 2a). In the clinical case of predictive oncology, based on scarce data, one should perform more sophisticated regression such as Kriging regression [8] to predict various scenarios and probabilities of cancer dynamics after treatment.

3.2 Model calibration

Before further experiments, it is necessary to calibrate the model to GT data and to estimate: initial values of dynamical variables, (C_0, P_0, Q_0, QP_0), and initial values of all the model parameters, which would give the best approximation of the *ground truth* data. These values will be later used as the first guess for creating the sub-models. Calibration has been performed by using the classical

Parameter name	Parameter value
λ_P	0.6801
K	160.14
$K_{Q_P P}$	0.0000001
K_{PQ}	0.41737
γ_P	5.74025
γ_Q	1.343
δ_{QP}	0.6782
KDE	0.09513
C_0	0
P_0	4.7279
Q_0	48.5147
QP_0	0

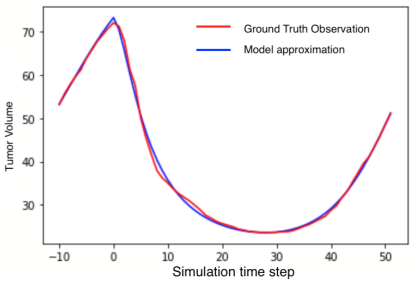


Fig. 3: The cancer model parameters after its calibration to *ground truth* data (left hand side) and the fitting of the calibrated model to *ground truth* for the first 60 time steps (right hand side).

$(\mu + \lambda)$ genetic algorithm (GA) for single-objective optimization. The fitness function has been defined as the root mean square error (RMSE) calculated for the *ground truth data* and tumor dynamics predicted for a given individual. GA has been executed in two phases. In the first one, the genotype of every individual consists of four double-float values representing the initial solutions C_0, P_0, Q_0, QP_0 whereas in the second stage the genotype of every individual

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

consists of eight double-float values representing the parameters of the tumor growth model (i.e., $\lambda_p, K, K_{Q_PP}, K_{PQ}, \gamma_P, \gamma_Q, \delta_{QP}, KDE$). In the experiments, the implementation of GA provided by the pyMOO[1] library has been applied. The calculated values of model parameters alongside the visualization of the tumor growth modeled by the calibrated model during the first 60 time steps are presented in Figure 3.

4 Experiments

The model of tumor dynamics used as the test case (see equation 9) has been selected because: (1) the computational time required for its simulation is reasonably short allowing for numerous repetitions needed in data assimilation and parameters matching procedure; (2) though the model is simple it is non-trivial in terms of the number of parameters; (3) the dynamics of the model can be disturbed anytime by administrating a new treatment what makes its dynamics unpredictable; (4) the model is realistic and can be considerably extended [11,12,21] making it computationally more demanding, what justifies using the supermodeling for its implementation and the use in clinical practice. The experimental

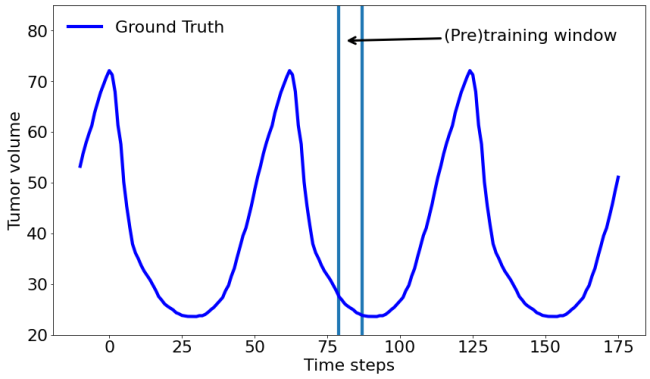


Fig. 4: Tumor dynamics in three drug administration cycles used as the *ground truth* data.

verification is focused on the analysis of prediction quality depending on the combination of: (1) the supermodel ensembling technique (*nudging*, *assimilation*, *weighted*), (2) the data assimilation and optimization algorithm (ABC-SMC [24], 3DVAR [23], classical derivative free algorithm, i.e., simplex method [9]) and (3) sub-model coupling scheme (i.e., dense or sparse sub-models coupling). For the purposes of this paper, i.e, demonstration of the usefulness and superiority of the supermodeling as a meta-procedure for data assimilation, we have selected a bit unrealistic tumor dynamics case shown in Figure 4. An oncologist who uses the predictive tool for planning the anti-cancer therapy would like to check all

[1] https://pymoo.org/algorithms/index.html

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

therapy scenarios depending on the tumor behavior. Usually, based on data from the beginning of the therapy, such as in Figure 2, they simulate further stages of tumor development planning the moments the following drug dose should be administrated to obtain the optimal therapeutic result. For sure, the oncologist would never allow for such an extensive tumor regrowth like that shown in Figure 4, moreover, he would use all the data from the beginning of therapy administration (not just a fragment which is shown in Figure 4) for further prediction. However, in our tests, we would like to study what accuracy can we expect for not only the "forward" but also the "backward" predictions. The latter allows for the insight into tumor past what could be very valuable for anti-cancer therapy and for inference of the reasons of the tumor evolution. Taking for our test GT data from the middle of the anti-cancer therapy, and having more complicated "backward" dynamics, we can better estimate the accuracy of the prediction methods we examined. The parameter values presented in

Table 1: Top-level experimentation parameter values

Parameter name	**Single model**	**Assimilated super model**	**Weighted super model**	**Nudged super model**
Evaluation budget	(30,50,70)	(30,50,70)	(30,50,70)	(30,50,70)
Pretraining ratio	N/A	40%	40%	40%
(Pre)training window range	[80 : 86]	[80 : 86]	[80 : 86]	[80 : 86]
Initial parameter perturbation	40%	40%	40%	40%
Number of submodels	N/A	4	4	4

Table 1 have been taken arbitrarily based on the preliminary experiments. The *evaluation budget* means the number of evaluations of tumor volume P^* (see equation 9) in the subsequent timesteps. In the experiments, the budgets are 30, 50 and 70, respectively. The *pre-training ratio* is the ratio of computational time needed for pretraining of the sub-models (in parallel) to the total training time, i.e., the evaluation budget considered (with training the supermodel coupling coefficients). We have assumed that pre-training is independent and thus can be run in parallel. In the following experiments, this value is set to 40% of the total *evaluation budget*. Initial perturbation means the deviation of the tumor growth model (i.e. λ_p, K, K_{qpp}, K_{pq}, γ_p, γ_q, δ_{qp} and KDE) from their reference values, where the reference values are collected in Figure 3. The parameters of the assimilation algorithms have been taken as provided by default in the pyABC[2] and ADAO[3] libraries.

4.1 Results

In Tables 2, 3, 4, 5 and 6 the average tumor dynamics percentage prediction errors are presented. The results in individual tables differ with tumor model dynamic variable the sub-models have been coupled by. The best tumor dy-

[2] https://pyabc.readthedocs.io/en/latest/
[3] https://docs.salome-platform.org/latest/gui/ADAO/en/index.html#

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

Table 2: Average tumor dynamics prediction errors. Coupling variable: P. (Results are in percentages.)

	Simplex			3DVAR			ABC-SMC		
	30	50	70	30	50	70	30	50	70
Single model	83,64	67,29	58,28	35,40	30,01	22,92	67,29	58,28	46,44
Assimilated supermo.	55,72	45,11	42,71	22,40	15,64	**13,01**	45,11	42,71	34,53
Weighted supermodel	49,51	42,12	39,31	22,23	16,61	16,54	42,12	39,31	29,78
nudged supermodel	30,28	29,66	27,08	22,15	16,96	15,43	30,28	27,08	21,10

Table 3: Average tumor dynamics prediction errors. Coupling variable Q. (Results are in percentages.)

	Simplex			3DVAR			ABC-SMC		
	30	50	70	30	50	70	30	50	70
Single model	78,46	49,57	39,07	40,88	33,22	24,84	57,50	57,09	47,47
Assimilated supermo.	50,04	40,81	34,30	24,94	23,72	19,12	42,19	38,45	26,32
Weighted supermodel	34,93	34,33	29,93	22,75	22,10	17,32	51,12	35,91	28,09
nudged supermodel	41,87	24,12	23,84	18,35	17,63	**12,75**	39,96	31,70	24,58

Table 4: Average tumor dynamics prediction errors. Coupling variable: QP.(Results are in percentages.)

	Simplex			3DVAR			ABC-SMC		
	30	50	70	30	50	70	30	50	70
Single model	56,38	42,96	38,44	43,84	24,10	21,97	72,41	62,04	49,51
Assimilated supermo.	37,70	30,51	25,27	21,92	19,54	16,01	44,95	40,47	34,35
Weighted supermodel	34,81	28,81	25,07	21,84	18,72	**15,93**	43,01	41,94	32,70
nudged supermodel	37,74	30,54	25,32	21,94	19,56	16,03	44,95	40,46	34,35

Table 5: Average tumor dynamics prediction errors. Coupling factor: C. (Results are in percentages.)

	Simplex			3DVAR			ABC-SMC		
	30	50	70	30	50	70	30	50	70
Single model	58,65	51,11	41,59	30,42	23,06	19,34	57,74	51,28	42,03
Assimilated supermo.	40,57	33,60	23,03	26,04	20,41	16,01	42,63	36,45	29,01
Weighted supermodel	48,73	32,01	24,24	27,59	22,75	**15,71**	42,91	37,60	26,38
nudged supermodel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Table 6: Average tumor dynamics prediction errors. Coupling variables: P, Q, QP, C. (Results are in percentages.)

	Simplex			3DVAR			ABC-SMC		
	30	50	70	30	50	70	30	50	70
Single model	58,83	51,91	46,21	31,81	26,03	19,34	65,14	56,14	33,95
Assimilated supermo.	35,54	31,73	24,45	25,72	21,16	19,12	49,5	34,19	25,96
Weighted supermodel	30,64	28,24	22,06	19,66	17,12	14,68	41,87	29,67	23,98
nudged supermodel	10,95	9,41	8,69	8,82	8,78	**6,22**	14,12	9,52	9,35

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

namics predictions produced by supermodeling approaches are as follows:

- when coupled by P – the nudged supermodel with 3DVAR (13,01%),
- when coupled by Q – the nudged supermodel with 3DVAR (12,75%),
- when coupled by QP – the weighted supermodel with 3DVAR (15,93%),
- when coupled by C – the weighted supermodel with 3DVAR (15,71%),
- when coupled by all the dynamic variables – the nudged supermodel with 3DVAR (6,22%).

For comparison, the best single-model predictions have been obtained with 3DVAR as a data assimilation algorithm and the average prediction error was: $21,68\%$. In Figure 5 we present the predictions and the accuracy of all compared

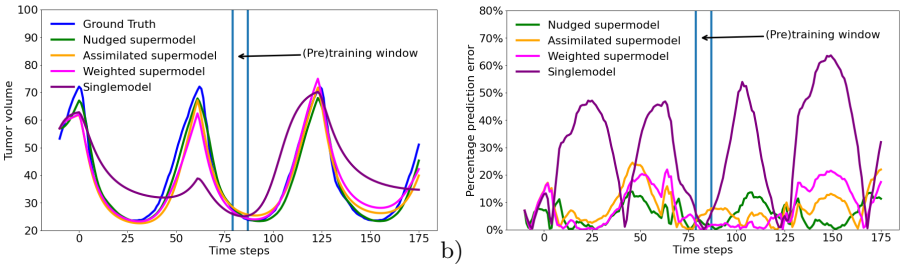


Fig. 5: Tumor dynamics prediction obtained by the single and the supermodels $Left$, and $Right$: the errors comparison

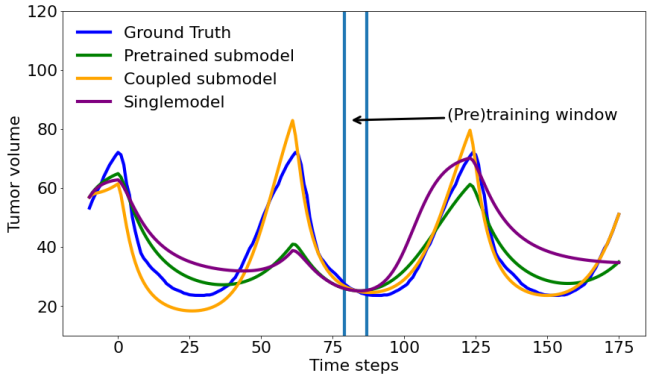


Fig. 6: Tumor dynamics prediction obtained by a pretrained sub-model (green line), the fully trained single-model (purple line) and one of the sub-models from the supermodel (orange line). The supermodel response is the average of the responses of all its member sub-models.

supermodeling schemes and the fully trained single-model. As one can see, the obtained prediction inaccuracies clearly demonstrates the advantage of the supermodeling approach. Thus, in the case of the single-model (the purple line in

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

Figure 6) if the DA process of parameters matching is not able to assimilate the model to reality with adequate accuracy, there is no mechanism that would be able to improve the prediction. While for supermodeling the situation is different. Even if the pre-training process assimilates the sub-model parameters very poorly (see the green line in Figure 6), what gives much worse predictions than the well-trained single-model does, then during the coupling of a sub-model with the other ones, it can be 'corrected' and synchronized with the others. Consequently, the supermodel ensemble, as the average of all the sub-models, is able to produce much better predictions (see the orange line in Figure 6) than the fully trained single-model.

5 Conclusions and future work

In the paper, the dependency of the supermodeling prediction quality on the supermodel ensembling method, the baseline data assimilation algorithm, and the (pre)training budget is analyzed. The combinations of three different supermodeling approaches, i.e., *nudged*, *assimilated* and *weighted* supermodels with three different data assimilation techniques (ABC-SMC, 3DVAR and SIMPLEX methods) have been analyzed. The tumor dynamics model has been used as the test case and its prediction accuracy has been compared to the generated *ground truth* evolution. The conclusions coming from the presented research are as follows: (1) All the supermodels clearly outperform the single-model accuracy in all the time budgets taken into consideration. The 3DVAR data assimilation algorithm, used as a baseline DA method, gives the highest accuracy. (2) For a single-coupling between the sub-models (only one dynamic variable is used for sub-models coupling) the accuracy of all the supermodel training methods is very similar and does not depend significantly on which specific dynamic variable is used to couple the sub-models. However, for dense connection (the sub-models are interconnected by all the dynamic variables) the average prediction error can be reduced by half and the *nudging* training scheme yields clearly the lowest prediction error. (3) The difference between the best supermodel and single model predictions is high. One should be aware that the more complex model the greater can be the difference. What is important, in many cases even a 1% difference in the prediction error (e.g. in climatology), may result in catastrophic consequences. The same one can expect in exploiting complex 3D tumor models in planning anti-cancer therapy. The perspective of the future works includes: (1) verification of the experimental results applying more complex and realistic tumor models than the test case used here, such as [11,12,21]; (2) applying another baseline data assimilation algorithms as 4DVAR, Kalman filters, etc.; (30 verification another supermodel ensembling schemes such as the "cross-pollination and pruning" technique proposed by Schevenhoven and Selten in [17,18].

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

6 Acknowledgement

The Authors are thankful for support from the National Science Centre, Poland grant no.
2016/21/B/ST6/01539 and the funds assigned to AGH University of Science and Technology by the Polish Ministry of Science and Higher Education. This research was supported in part by PLGrid Infrastructure and ACK CYFRONET Krakow. We also thank professor dr Greg Duane, University of Bergen, for his helpful comments and advice.

References

1. Bouttier B., Courtier P.: Data assimilation concepts and methods, Meteorological Training Course Lecture Series, European Centre for Medium-Range Weather Forecasts, 1999
2. Dzwinel, W., Kłusek, A., Vasilyev, O.V.: Supermodeling in simulation of melanoma progression, International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA, 999–1010
3. Evensen G.: The Ensemble Kalman filter: theoretical formulation and practical implementation, Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, European Centre for Medium-Range Weather Forecasts, 8–12 September 2003
4. Glover, F. W.: Tabu search in Encyclopedia of Operations Research and Management Science, Springer US, 1537–1544, Boston, MA, 2013
5. Glover, F.: Tabu Search part I, ORSA Journal on Computing, 1(3), 190–206, 1989
6. Glover, F.: Tabu Search part II, ORSA Journal on Computing, 2(1), 4–32, 1990
7. Hazeleger, W., van den Hurk, B., Min, E.: Tales of future weather. Nature Climate Change 5, 107–113, (2015)
8. Hengl, T., Heuvelink, G. B., Rossiter, D. G.: About regression-kriging: From equations to case studies. Computers and geosciences, 33(10), 1301-1315, 2007
9. Karloff, E.: The simplex algorithm, Linear Programming, 23–47, Boston, 1991
10. Kirtman, B.P., Min, D., Schopf, P. S., Schneider, E. K.: A New approach for coupled GCM sensitivity studies, The Center for Ocean-Land-Atmosphere Studies, Technical Report, 2003
11. Klusek, A., Los, M., Paszynski, M., Dzwinel, W.: Efficient model of tumor dynamics simulated in multi-GPU environment. The International Journal of High Performance Computing Applications, 33(3), 489-506., 2019
12. Łoś, M., Paszyński, M., Kłusek, A., Dzwinel, W.: Application of fast isogeometric L2 projection solver for tumor growth simulations, Computer Methods in Applied Mechanics and Engineering, 306, 1257–1269, 2017
13. Powell, M.J.D.: An efficient method for finding the minimum of a function of several variables without calculating derivatives, Computer Journal, 7(2), 1964, Oxford Academic, 155-162
14. Powell, M.J.D.: A direct search optimization method that models the objective and constraint functions by linear interpolation, Advances in Optimization and Numerical Analysis, Springer, Netherlands, 51–67, 1994
15. Qing, A.: Fundamentals of different evolution in differential evolution, Fundamentals and Applications in Electrical Engineering, 41–60, Wiley-IEEE Press, September, 2009

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28

16. Ribba, B., Kaloshi, et al.: A Tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy, American Association for Cancer Research, Clinical Cancer Research, 18(18), (2012), 5071–5080
17. Schevenhoven, F.: Training of supermodels-in the context of weather and climate forecasting. PhD thesis, University of Bergen, Norway (2021).
18. Schevenhoven, F. J., Selten, F. M.: An efficient training scheme for supermodels, Earth System Dynamics,8(2), 429–437, 2017
19. Sendera, M., Duane, G. S., Dzwinel, W.: Supermodeling: the next level of abstraction in the use of data assimilation, in Proceedings of the International Conference on Computational Science – ICCS 2020, 133–147
20. Siwik, L., Los, M., Klusek, A., Dzwinel, W., Paszyński, M.: Tuning Two-Dimensional Tumor Growth Simulations, Proceedings of 2018 Summer Computer Simulation Conference SCSC 2018, July 9-12, 2018, Bordeoaux, France, pp. 8, ACM Conference Series, 2018
21. Siwik, L., Los, M., Klusek, A., Pingali, K., Dzwinel, W., Paszynski, M.: Supermodeling of tumor dynamics with parallel isogeometric analysis solver. arXiv preprint arXiv:1912.12836., 2019
22. Suganthan, P. N.: Differential evolution algorithm: recent advances, Theory and Practice of Natural Computing, 30–46, Springer Berlin Heidelberg, 2012
23. Talagrand, O.: Assimilation of observations, an introduction, Journal of the Meteorological Society of Japan (Special Issue on Data Assimilation in Meteorology and Oceanography: Theory and Practice), 75(1B), 191–209, 1997
24. Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P.H.: Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, Journal of the Royal Society Interface, 6(31), 187–202
25. Weigel, A. P., Liniger, M. A., Appenzeller, C.: Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?, Quarterly Journal of the Royal Meteorological Society, 134(630), 241–260, 2008
26. Wiegerinck, W., Burgers, W., Selten, F.: On the Limit of large couplings and weighted averaged dynamics, Consensus and Synchronization in Complex Networks, 257–275, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2013
27. Wiegerinck, W., Mirchev, M., Burgers, W., Selten, F.: Supermodeling dynamics and learning mechanisms. Consensus and synchronization in complex networks (pp. 227-255). Springer, Berlin, Heidelberg.
28. Wiegerinck, W., Selten, F.M.: Attractor learning in synchronized chaotic systems in the presence of unresolved scales, Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(12) (2017),1269-2001
29. van den Berge, L. A., Selten, F. M., Wiegerinck, W., Duane, G. S.: A multi-model ensemble method that combines imperfect models through learning, Earth System Dynamics, 2:(1), 161–177, 2011
30. Yamanaka, A., Maeda, Y., Sasaki, K.: Ensemble Kalman filter-based data assimilation for three-dimensional multi-phase-field model: estimation of anisotropic grain boundary properties. Materials & Design, 165, 107577, 2019

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77964-1_28