SocultPaperV3

2024-05-19

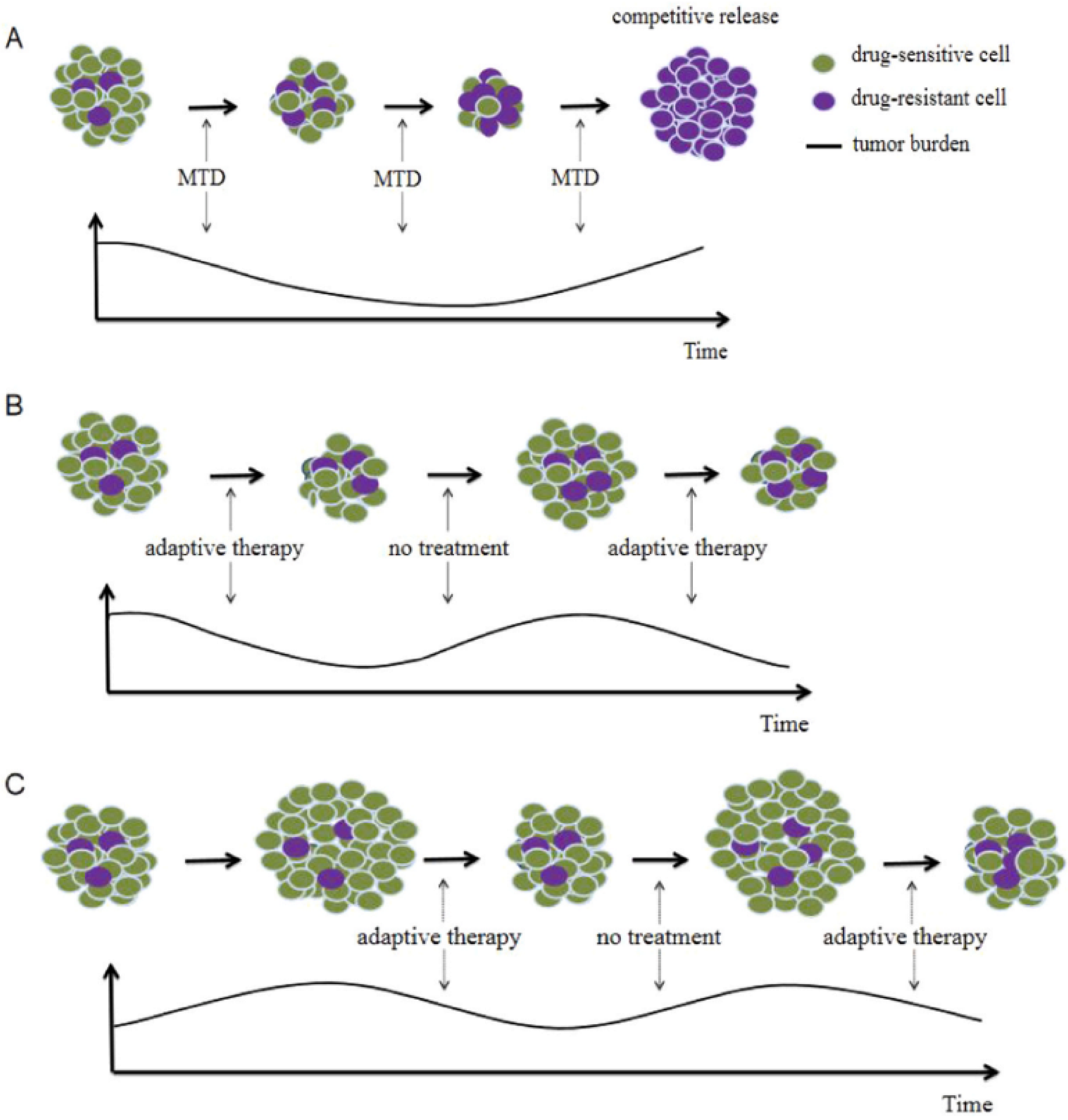
Introduction

What is the adapative cancer therapy

Worldwide, cancer is the cause of 1 out 6 deaths. Of these cancer deaths an estimated 90% are due to development of drug resistance (Bukowski, Kciuk, and Kontek 2020). While intial cancer treatment usually shows positive response in tumor burden, drug resistance develops due. To highlight the inefficincy of traditional approaches, (Staňková et al. 2019) models cancer treatment game theorict contest between a physician and a tumor, where physicians move on each round is to apply a certain treatment, and a tumor makes an adaption. While it is a "Stackelberg game", a game where one player is the leader (the physician) and another player is a follower (the tumor), is assymerty is rarely exploited. Instead of using their advantage to steer the evolutionary pressures placed on tumors, the physician lets the tumor not only adapt to the current round of the game but also to future rounds of the game. The advantage of leading the game is thereby lost. The authors aptly analogize the current practice:

> *"Consider cancer treatment as a rock-paper-scissors game in which almost all cells within the cancer play, for example, "paper." It is clearly advantageous for the treating physician to play "scissors." Yet, if the physician only plays "scissors," the cancer cells can evolve to the unbeatable resistance strategy of "rock.""* (Staňková et al. 2019)

Adaptive Therapy is a approach to cancer treatment based on controlling the intra-tumoral evolutionary dynamics. By leveraging that cancer cells can incur a fitness cost to evovling mechanisms that yield resistance to drugs. For example it has been shown that tumor cells can mutate to increased expression of PGP membrane pump, which uses ATP to move drugs out of the cell. While this makes cell more resistant to treatment, it also comes at metabolic cost. (Gatenby et al. 2009) found that PGP activity was the culprit of approximately 50% of cell metabolism. As a result, if resistant and non-resistant cells are competing for space and resources, drug-sensitive cells should over time outcompete resistant cells. Apative Therapy utilizes this darwinian competition to make cancer fight itself. Thus the dream-sceneario of this adaptive therapy is not to eradicate cancer, but instead make it a controllable chronic disease.



Initial results from pilot clinical trial on metastatic castrate-resistant prostate cancer patients are promising. Intitial results showing both less cumuluative dosages and longer survival in comparison similiar group of patients recieving standard care. Firstly, patients were only involved if they showed a substainal positive response to treatment. The trial has then been utilizing a range-bounded treatment rule. If the bloodmarker *Prostate-specific Antigen* (PSA), a proxy of tumor burden, increases back to pre-trial levels, treatment is applied until PSA drops to 50% of pre–trial levels (Zhang et al. 2017). The trial is expected to run until december 2024 (Center and Research Institute 2024).

Desired qualities of models

While the trial reported by (Zhang et al. 2017) only utilized a single drug, key researchers in researchers in Adapative Therapy has produced a review of the use of mathematical modelling in the field, and among other things, indentified that modelling multidrug treatments is a necessity of future models. Additionally, they argue that it is unlikely that any treatment approach can accomplish delaying the emergence of resistant cells, lower the tumor burden and minizmize the toxicity. Given that patients likely differ in their ability to tolerate tumor burden and the toxicity of drugs and the evolutioanry dynamics of the cancers that they carry differ, ,odels would ideally therefore have to trade-off each these factors given a specific patient. This will require fitting data on invidual patients. Lotke-volterra models have been fitted frequently.

biological factors Another important aspect of modelling, is illumanitng the actual comppetive disadvantage that resistant cells are. While larger tumor sizes are thought to increase the suppresion of resistant cells, the dynamics are likely much more complex. Factors such as the spatial configuration of actual cells and the range of the molecular influence they yield of each other. It is also crucial that these actually incur a fitness cost, however other authiors argue that adaptive therapy might still be able to delay time-to-progression if this is not the case. Comptettion isn't neccesarily strong if the tumor isn't at carrying capacity.

** CONTROLLING THE RESPONSE TO TREATMENT IS PARAMOUNT. MOUNTAIN CAR PROBLEM**

Gene-expression has been shown to change in cells as a result of treatment, which further complicates modelling the adaptive therapy as case of resistant vs non resistant cells. Phenotypic plasticity should therefore also be accounted for. Another biological factor that should be accounted for is sourrind tissue. For example, prostate cancer cells in bone can utilzied such as the transforming growth factor β can accelerate the proliferation of cancer cells.

** You can keep throwing layers at POMDPs**

The efficay adapative therapy depends strongly on initial resistance rates, and deciding to opt for control strategy such as adpative therapy would be beneficial if made early. THis however necessitates predicting what patients would respond better to adaptive therapy and who benefit better from the other treatment protocols, such as standard maximum tolerable dose protocol.

How to construct dosing protocols High tumor burdens might also come with other costs, such as increasing the risk of new metastates or simply by the fact that more cells increase the chance total amount of mutations happening. Adapative therapy also depends on frequent monitoring, and could benefit from the use of different testing protocols.

Robustness to changes to in plans due to machine failure or tother practical constraints.

There is need to include multi-drug treatment protocols in adaptive therapy, but the number of possible permtuations at each point in time grows extremely fast when more treatment options are introduced. A potential solution to this problem is constructing a treatment protocol that steers the tumor in cycle, so that the conditions at the start of one treatment block is indentical to how conditions of the prior block. In principle only one block would have to be designed then. ** this is a non-steady-state equlibirum **

Constructing real time prediciton Prediciting individual patient repsonses real-time would greatly enhance adaptive therapy since dose modulation could be indivdualized further and evolutionary dynamics controllable with more precision. Using relevant biomarkers would be crucial in this regard ** all observable consequences also generate information ** Any chosen model must be able to be calibrated and validated before hand. When to time the collection of biomarks also seems to be in issue, since the prostate trial found that treatment would at times overshoot, since biomarkers were collected too late, and the PSA levels were dropped well-below 50%. This likely leads to poorer clinical performance. The link between biomarkers and actual tumor progression is not certain neither, meaning that decidnig when to treat directly on the basis of a biomarker might not be optimal.

A key issue going forward is rethiking how data on patients is collected. It will be crucial to collect data not only to detect progression, but to collect data that will be usefull for future decisions too. ** this is EFE** Quantifying the unceartinty in the models belief and the consequences of its suggested actions is also paramount.

What are POMDPs in general

Partially Observable Markov Decision Processes (POMPD) is class of controller models that model and underlying markovian process, that in disctrete time and state enviroment, the next step of the system only depends on the current step. Crucially for partial orbservability is crucial facet of these models, and refers to the fact that these models don't directly observe the actual enviroment or markovian process, but instead only potentially noisy signals emitted by it while trying to manipulate the evivorment (Åström 1969). This allows these models to differiate an orbserved signal from what it "believes" about the enviorment and use a single reward function trade off unceartinty for achieveing a certain goal state (Kaelbling, Littman, and Cassandra 1998) while yielding bayes optimal beliefs. For example, if discretized a, POMDP could understand a PSA reading as noised signal of actual tumor state, and thus try to control to tumor state rather than the PSA reading. While POMDPs are typically difficult to solve analytically various approximate approaches exist.

Using active inference to solve pompds

One approximate solutions, of these is born out of the field of neuroscience litterature. The field which has come to be know as active inference suggests that the brain could by using a varitional approach. By minimizing two objective functions. *Free Energy* (EFE) as measure of model and past sensory inputs, and *Expected Free Energy* (EFE) which evaluates future courses of actions against a set preferred observations. Active Inference has been used to model psychopathology (Da Costa et al. 2020) but also applied to control scenearious such as the mountain-car problem (Friston, Daunizeau, and Kiebel 2009) and, albeit augmented with deep-learning, robotics control @(Çatal et al. 2020). These have been implemneted in MATLAB and recently in Python with the python package pymdp [(Heins et al. 2022a) and (Heins et al. 2022b).

Typically POMDPs are modelled for discrete state spaces. Mathematically they are described as a joint probability:

$$p(o, s, u; \phi)$$

where o are observations, s are hidden states, u are "control states" (states that an agent can influence) and ϕ are the hyperparameters of the model, such as α typiccaly used as 'inverse temperature' i.e. how deterministicly the model selects actions. By conditioning on certain observations, the POMDP is solvable by various approximations schemes for the optimal posterior beliefs over what the underlying hidden states are and what the optimal course of action is given a set of preferences. For example, a POMPD could model the joint probability of observing a certain amount of tumor burden in a patient, given some hidden states, such as the how resistancy of the underlying tumors, and whether a certain treatment was applied. Disregarding computability this can be done for any time horizon. Due to considering time and states discrete, this joint probability can be made tractable for a time horizons by factorizing the joint probability into the following categorical probability distributions:

- A likelihood model A. Ususally modelled as a set of arrays where each array corresponds to an observation modality describes the how observations map to a particular state. For example how one modality could be PSA readings, and its likelihood array describes how likely different test results are under different tumor burdens. It can both be modelled as a perfect signal of the tumor burden, ie the same tumor burden always gives the same test result, but it could also be implemented as noised signal. If multiple states are modelled, each likelihood modality is an array with a dimension corresponding to the number of possible observations of that modality, and for each hidden state another dimension is added with the same length as the number of possible hidden states. In this way, every combination of state can be mapped to an observation.
- A transition model B. It describes the probability of state transitioning to another on any particular timestep. This also encodes how actions are expected to influence transition probabilites. It could describe how a tumor is likely to evolve from one timestep to next depending whether treatments

is being applied. The transition model is usually coded as collection of three-dimensional arrays, a first dimension for the next state, a dimension for the current state, and a third dimension with the length of each action that would influence transition probabilities of the state. Modelling the transition probabilities would depend on whether treatment is being applied or not.

- A prior over preferred states C. A particularity of these models is that utility is specified in probabilities. This prior is set over certain observations, meaning that the model artificially expects to see certain observations. As a result, this drives the models behavior to act in ways that bring it close these expectations. At a first pass, describing a goal in terms of probabilities seems odd, but it could be considered a fundamental property of self-organizing systems. For example, to stay a live, I would have to spend time states where my tumor burden is low, even though in all likelihood, given enough time, I would come to develop cancer. Crucially, it allows for inferring what actions would bring about me spending time in states with low tumor burdens.
- A prior over initial states D. These are vectors of probabilities of intial beliefs. In continuation with example above, this component would specify how probable different levels of tumor burden before interacting with the patient.

The POMPD can be solved at any time step by minimizing two measures, *Variational Free Energy* (VFE) and *Expected Free Energy* (EFE). VFE is an upper bound on *surprisal,* and is approximated to infer the most likely current state given an observation, the likelihood model, and the prior over either initial states if no earlier orbsevations have been made. Otherwise the the precedding posterior over states is used as the current prior (Smith, Friston, and Whyte 2022). VFE scores how well the models representation aligns with a set of observations, and by using different approximate posteriors over states, choosing the approximate which minizimes VFE will approximate bayesian belief updating (Da Costa et al. 2020). EFE instead score, plans on how to act and their expected outcomes. EFE is used to construct a posterior distribution over what policy is most preferred. Since the model is equipped with C, the set of prior preferences over observations, the distinguishing between probable and prefered is difficult. This could however be considered a benefit, since it EFE is a composite of utility gain and information.

The aim of this paper

This potential benefit to adaptive therapy could be realized if adaptive therapy is modelled as a system where the tumor burden emits a noised signal. While controlling the tumor burden in short run is imperevative, truly successful strategies will have to also control future resistance dynamics. This task seems more challengig since resistance dynamics don't appear directly observable in clinical settings. Two obvious choices emerge, either implement a heuristic such which we expect to control the resistance dynamics well, or explicitly model the resistance dynamics despite paucity of real-time data on resistance. However, by applying a treatment and orbserving how the changes in tumor burden, the underlying resistance dynamics should be inferable. This paper invesitagets whether active inference can be used to gain traction on the second strategy.

Consider the following motivations for using Active Inference:

- Real-time data will be sparse in the clinical settings. Extracting the maximum amount of infomraiton for each data point could be pivotal. Certain implementations of Active inference purpotedly approximate optimal bayesian inference.
- Short-term exploiting tumor vulnerability is essential for keeping the patient alive, but for long term success, but controlling resistance dynamics is essential. A therapy plan will have to balance keeping the patient alive now against limiting its future options for treating. Since the degree of resistance is not directly
- Gaining information about how the resistance dyanimcs is also essential and this must be done through treating the tumor. Choosing whether to treat would therefore not only be a consideration of the tumor level, but also about how much wiser it would make us about learning the resistance dynamics.

- If the resitance level is only inferable through orbserving the treatment efficacy, the expected information gain from observing tumor the burden, i.e. testing, will differ between periods of treatment and non treatment. All else being equal, testing would therefore be more valuable during treatment.

An interesting corollary is that two patients could have the excatly the same tumor burden, but if the resitance dynamics is well modelled for one, treating could be poor choice. But the other patient could have excatly the have the same tumor burden, applying treatment could be the optimal move simpllearn about the resistance dynamics of their cancer. Due its approximations of bayesian inference and its ability to make utility information tradeoffs when deciding on actions, active inference seems to be promising paradigm planning adaptive therapy. By constructing simulating a simplified disctrete enviroment based on adaptive therapy, this paper will attempt to adapt the active inference implmentation of POMDPs in pymdp with the goal of keeping a virtual patient alive for as long as possible by inferring and controlling the tumor level in the short run and the resiostance level in the long run.

Analysis

Simulated environment

In order to easily comply with the discrete state and time implementation typical of the active inference litterature, discetrete states and time were used. Each run of the enivorment were maximally 200 timesteps long. In each timestep, a model has to keep a virtual cancer patient alive. Only one treatment and testing type exists. This was inspired by the use of PSA as testing of choice, and Abiraterone as the only treatment in a pilot clinical trial (Zhang et al. 2017). In the simulation, a patient's tumor state determines whether they survive to the next timestep or not. Runs always begin with the tumor state at 0, but it increases with a fixed risk at each timestep. If the tumor state reaches 5, the simulation ends.

To avoid this, a model has decide when to apply treatment. Applying treatment can reduce the tumor state. But whether a treatment is succefully reduces the tumor state depends on a underlying *resistance state*. The resistance state also begins at state 0 out of 5. When the resistance is low, the chance of treatment succeeding is high and vice versa. But for each round of treatment, the resistance state has a fixed risk of increasing. The resistance state can decrease if treatment is withdrawn, but the probability of the resistance state decreases depends on the tumor level. At high tumor levels, the resistance state is more likely to decrease. This is based on the work of (Hansen and Read 2020) who suggests that larger tumor sizes should generate more competitive pressure on resistant-cells.

All in all this means that a model has to balance the not tumor growing out of control, which would "kill the patient", against not painting itself into a corner by applying treatment too frequently. In order to successfully long-term manage the disease, a model will therefore also have to manage the resistance state. To mimic, the dynamics of the pilot trial with mCRPC, only the tumor state can be observed. The resistance state will therefore have to be inferred through how the tumor state changes when treatment is applied.

Modifications to POMPD scheme

Typically in active inference implementations of POMPDs, hidden state factors such as resistance and tumor states do typically not affect each other. Instead the interactions of different state factors are typically modelled as leading to different observations. This means that they coded in the A-array likelihood mappings, and in C prior preference distributions.

This setup was deemed inadequate for the current experiment. Instead a model should be able to infer an causally upstream state-factor, such as the resistance state, through down-stream partially orbsevable state-fators, i.e. the tumor state. This approach should allow further complexify models, and add more causal nodes occluded causal nodes.

Specifically, modifications to the transition probabilties in the generative model were made. Usually, three dimensional “B-tensors” describe expected transition probabilties within a state factor: one dimension the current state, one dimension for what ever action is chosen and third for the resulting state. For the present project another dimension was added, this dimension corresponds to the state of another state factor. Concretely, this meant that the tumor state factor had a fourth dimension corresponding to tumor transition probabilities for each resistance state. The expected transition probabilities could then be estimated by matrix multiplication between the B-tensor for the tumor state factor and the expected resistance level at the corresponding time point. One can imagine that instead of having only one B-tensor for the transition probabilities of the tumor state, one was created for each resistance level. The change simply amounts to taking a average of all these tensors weighted by the expected probabilities for each resistance state. This strategy was repeated for the resistance level, since decreases inthe resistance level depended on the tumor state. While rendering much of the pymdp functianality immediately unsuable, this strategy should be able to model arbitrarily complex dependencies between states. Hopefully, a robust implementation can be designed without overhauling pymd. It should be noted that this modification means that it is longer meaningless which order beliefs about states are evaluated and that this must be specified. For example for the present simulations, the effect of resistance level on the treatment efficacy was evaluated first, since this was evaluated first at at each step of the simulation.

Exploratory Simulation

An exploratory simulation, where a single run of POMDP controlling treatments and testing was run to investigate the feasibility of using the modified POMDP scheme. This model could had to choose whether test and treat. It would always observe a perfect signal of whether the patients is alive, whether model is testing and whether treating was being applied. If the model decided to test, it would receive a signal of the tumor state. Its prior preferences were heavily against observing a dead patient, somewhat against treating and little against treating. This was done to simulate a cost to both treating and testing, This means that the model will have to balance the utility gained and lost “cost” of all these actions” while considering the potential information gain of each choice. The model was further handicapped by the noising the tumor signal. This means that the the resistance state, which must be inferred through the tumor signal is doubly obfuscated. However the likelihood mapping between mapped the expected noise in the signal of the tumor signal, and it perfectly knew the transition probabilities of the environment. The model was given uniform priors over intial states, meaning it had no knowledge at beginning of each run.

Learning underlying hidden state

The simulation run manages to keep the patient alive for 68 timesteps and chose to test on 55 of these (see Fig. 1)

Only the tumor level is observable to the POMDP. It must combine its knowledge about how resistance level likely increases after applying treatment, and how the tumor level responds to treatment depenping on the the resistance level. While the model far from perfectly knows the resistance level. Model beliefs for the entire run is plotted.

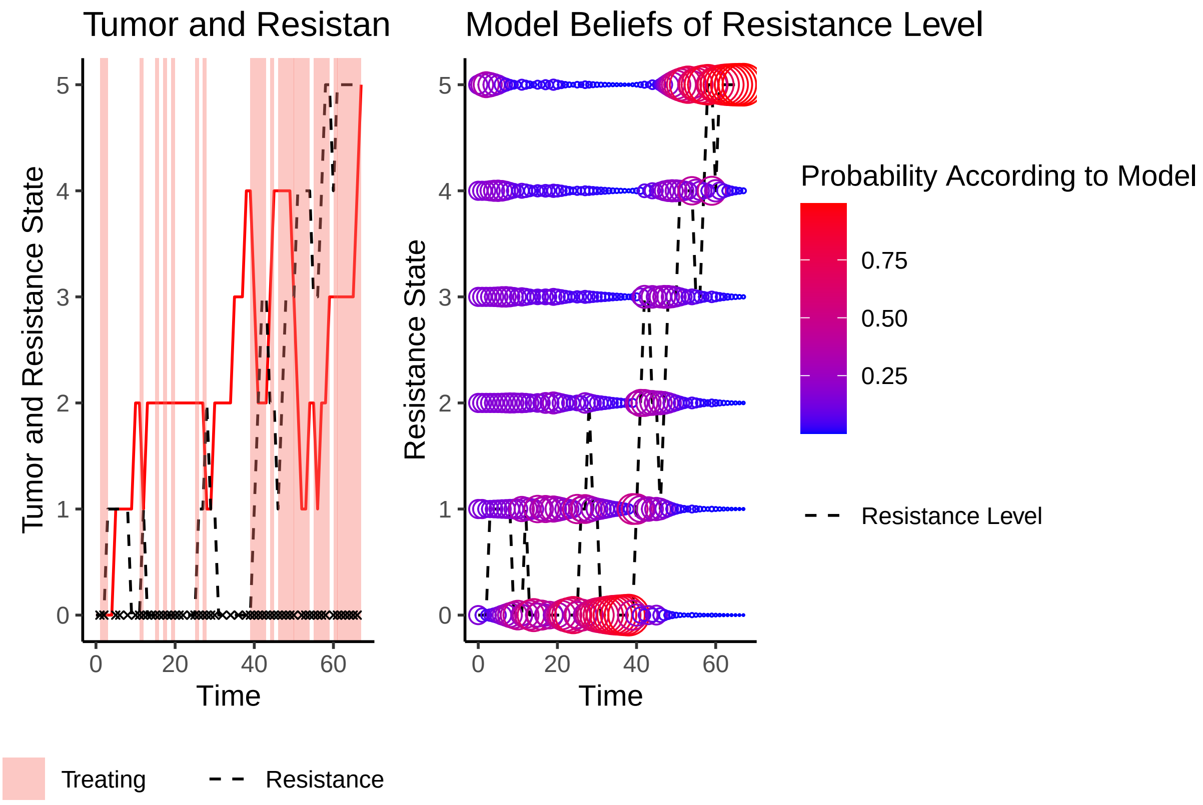
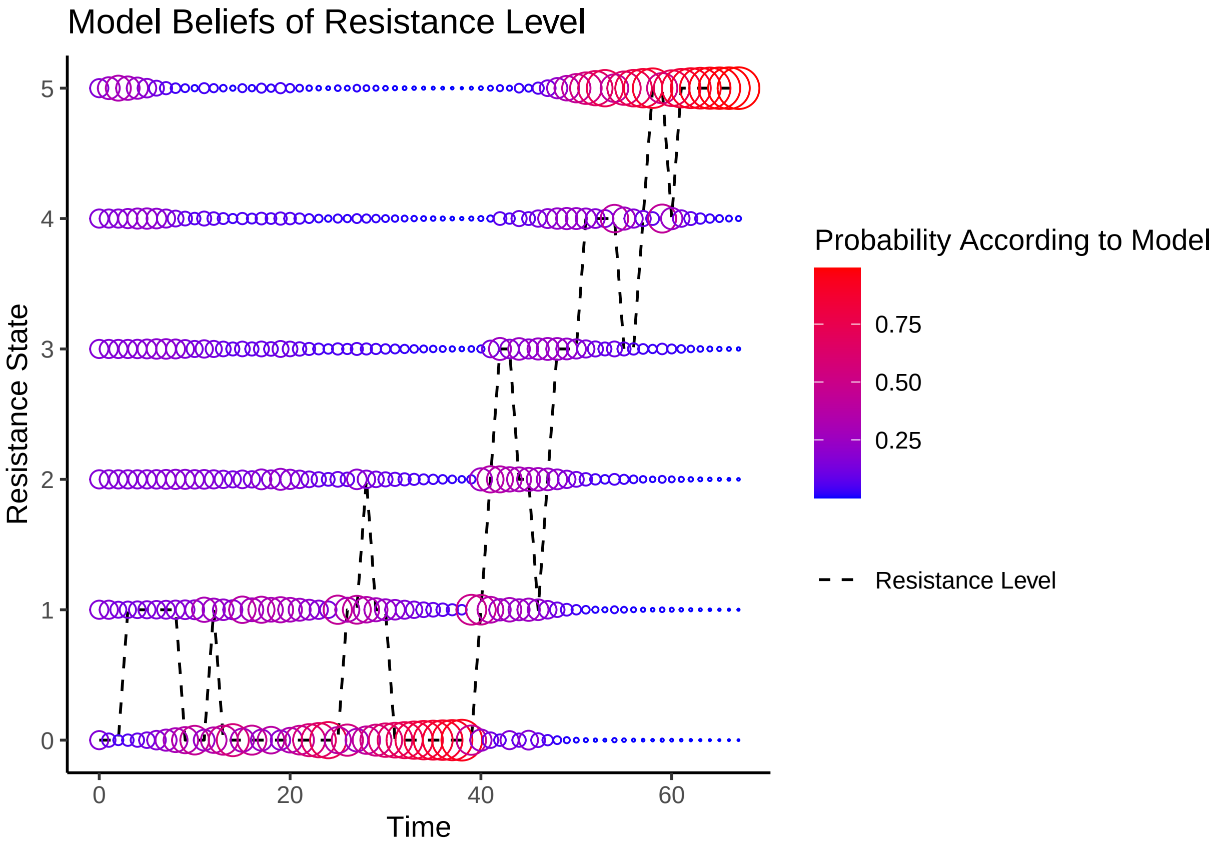
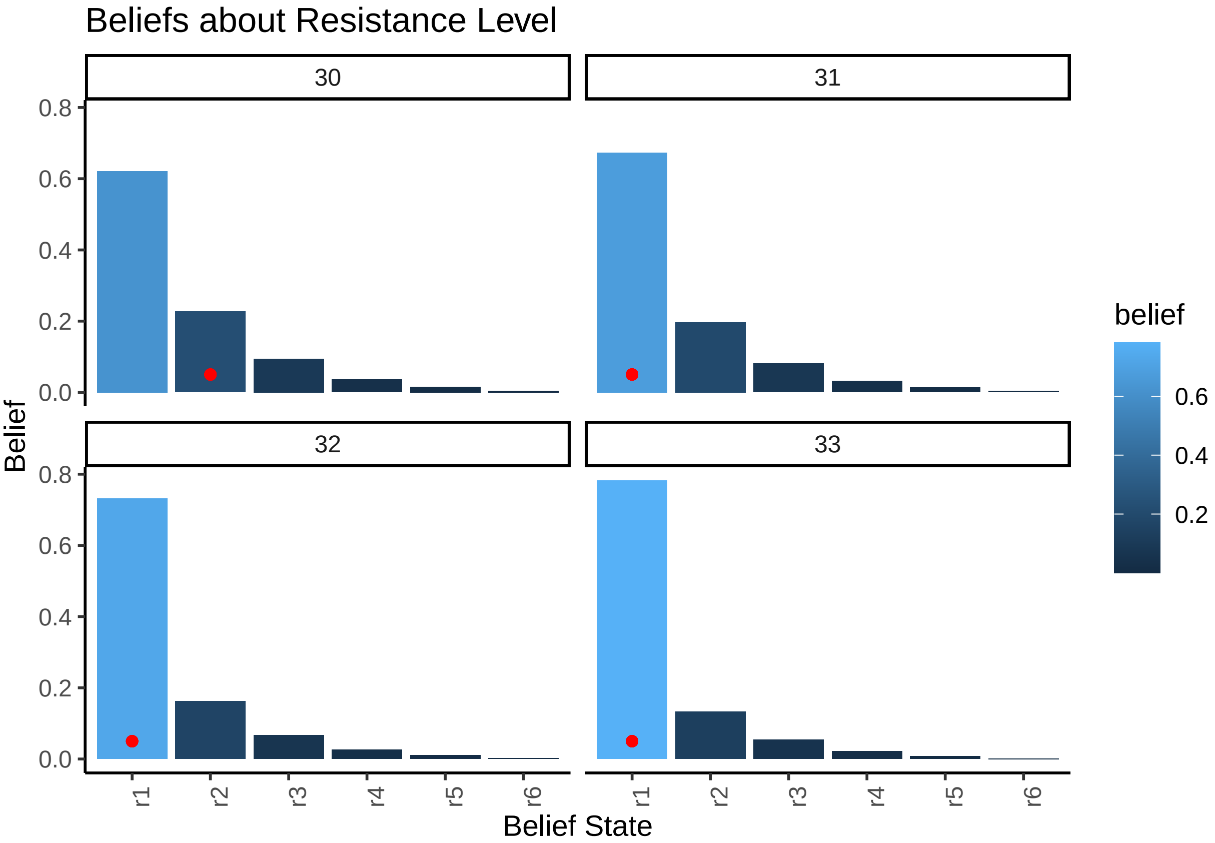


Figure 1: Figure 1. The left panel shows a simulation run for explotary analysis. Th he



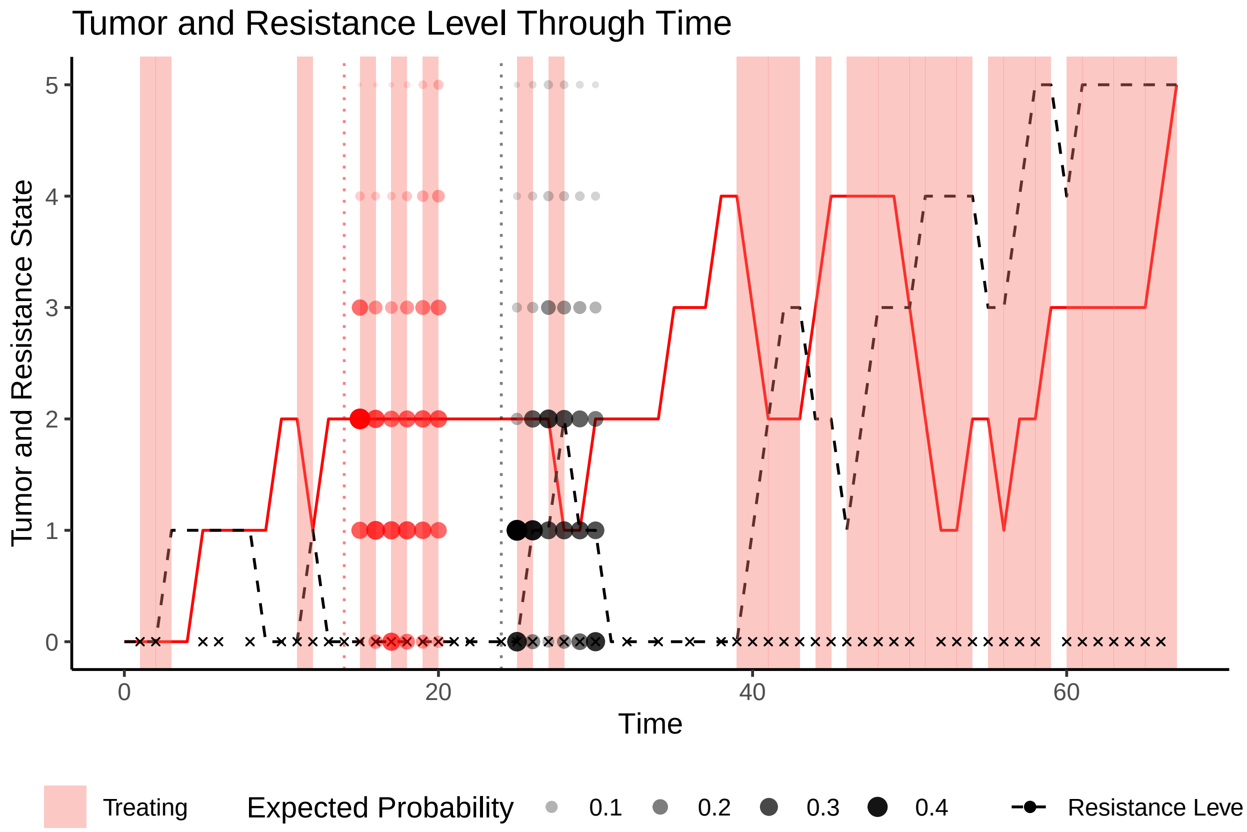
The model begins fairly agnostic. Given that the model was initialized with a uniform prior over resistance states this makes sense. Thorought out the course of the simulation it then finds loweer values of resistance state more likely.The POMDP structure is capable of inferring an underlying hidden state: the resistance state. Even though this not directly observable, this can be inferred to how the tumor state responds to treatment. Applying treatment for longer time without any beneficial effect would suggest that the resistance level is high, while immediately observing that the tumor level decreases would suggest that the resistance level is low. At each timestep the POMDPs perform infer the most likely state of every state factor, given their current observation and prior beliefs. Through custom changes to the 'get_expected_states' function in PYMDP that allowed the models to consider how one hidden state factor (resistance factor) would influence another (the tumor factor).



The above plot shows the strength of beliefs in t probabilities for a subset of timesteps [30 - 34], and the red dot show the actual the resistance levels. A time progresses the resistance level increases, and the model adjusts its beliefs. While the tumor level doesn't increase durings this time_period, this would also signal the model that the resistance level is low. This is the case since the enviroment always has change of increasing the tumor state, but such an increase would be negated by succesful round of treatment.

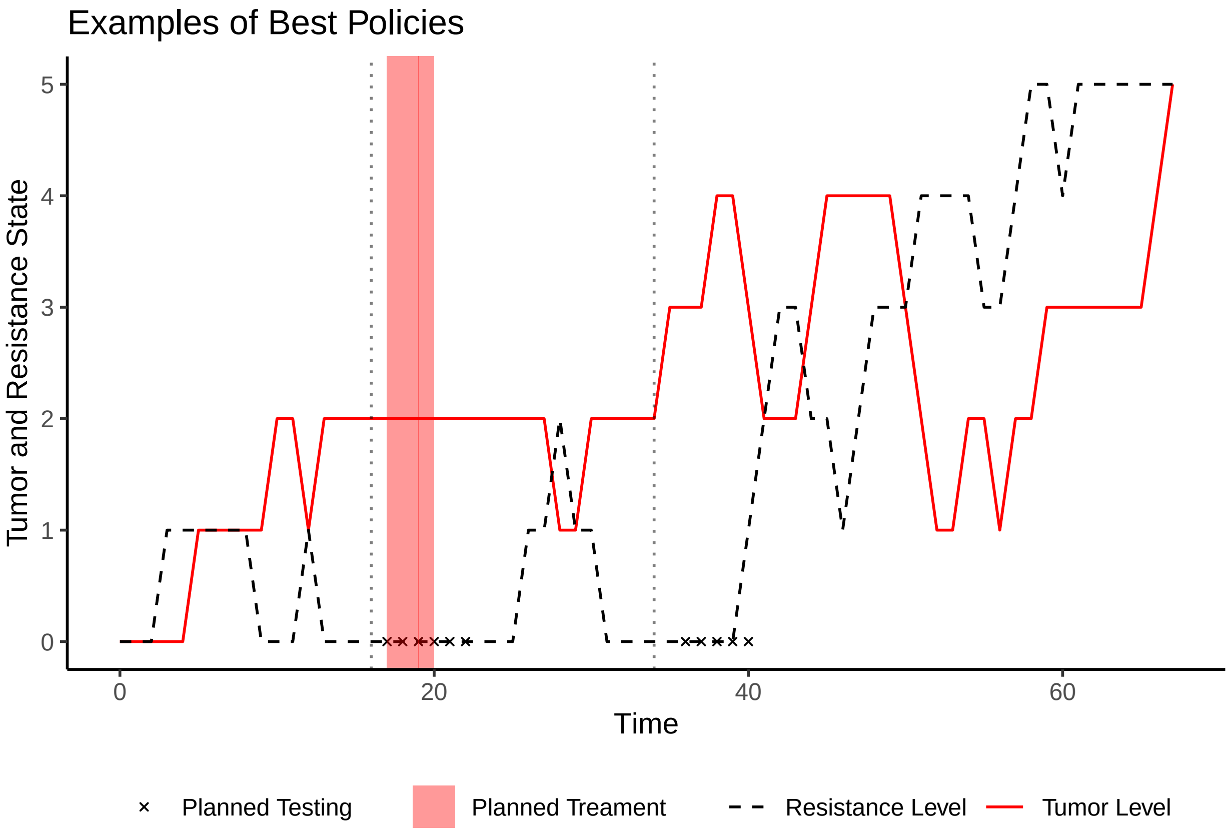
Beliefs, current/future and uncertainty is accessible

This current simulated model used a slightly different implementation than those compared to performance of rangebounded therapy planning. While the those models had longer policy horizons, they didn't consider the entire space possible actions. A model with a shorter policy horizon that would instead search every single possible action four steps is plotted to investigate the structure of decision-making by the POMDP model. The enviroment was also more unceartain. Instead of featuring a 1-to-1 mapping of the observations of the tumor to actual tumor state, it recieved a noised signal. It must therefore. The beliefs about how tumor state will evolve as a consequence of the most promising policy at time step 14 is plotted, and the expectation of resistance states at t=24 is plotted too.

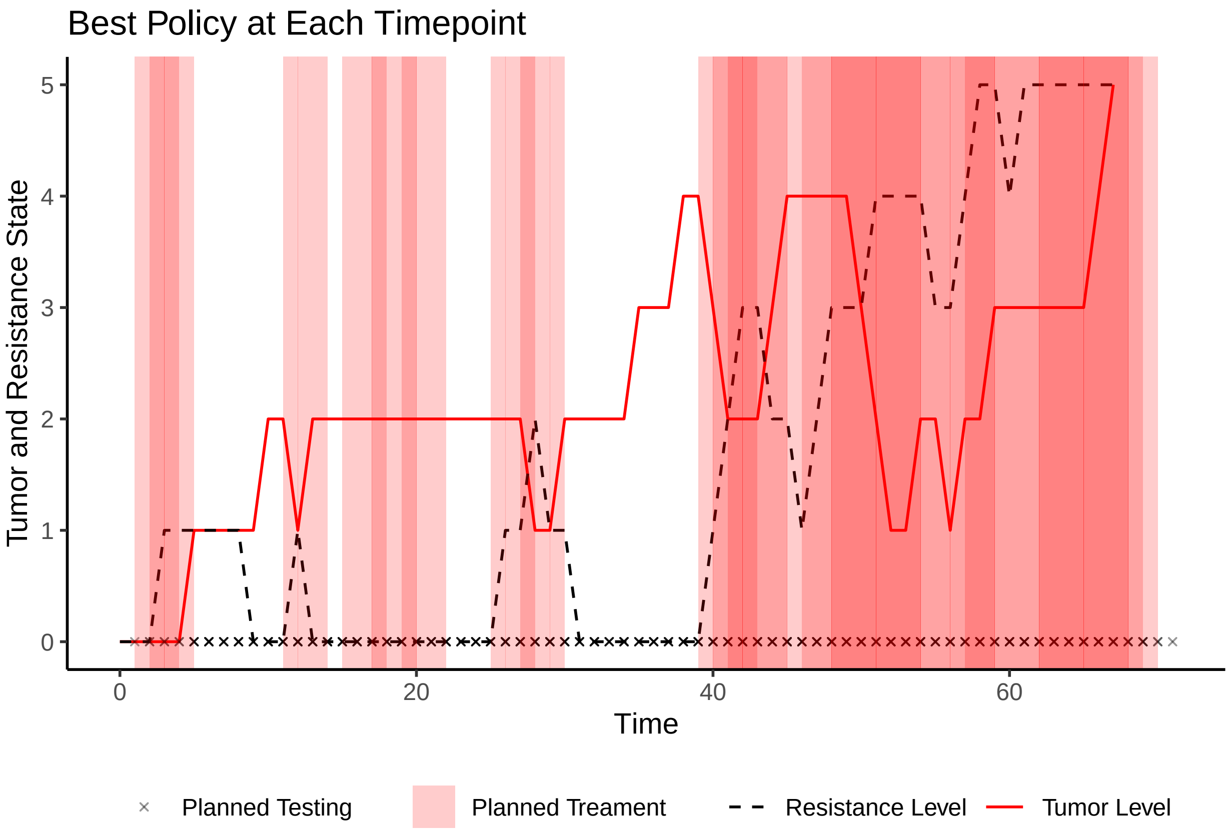


Dotted vertical lines indicated the time-point for which the evaluation of expected states are extracted.

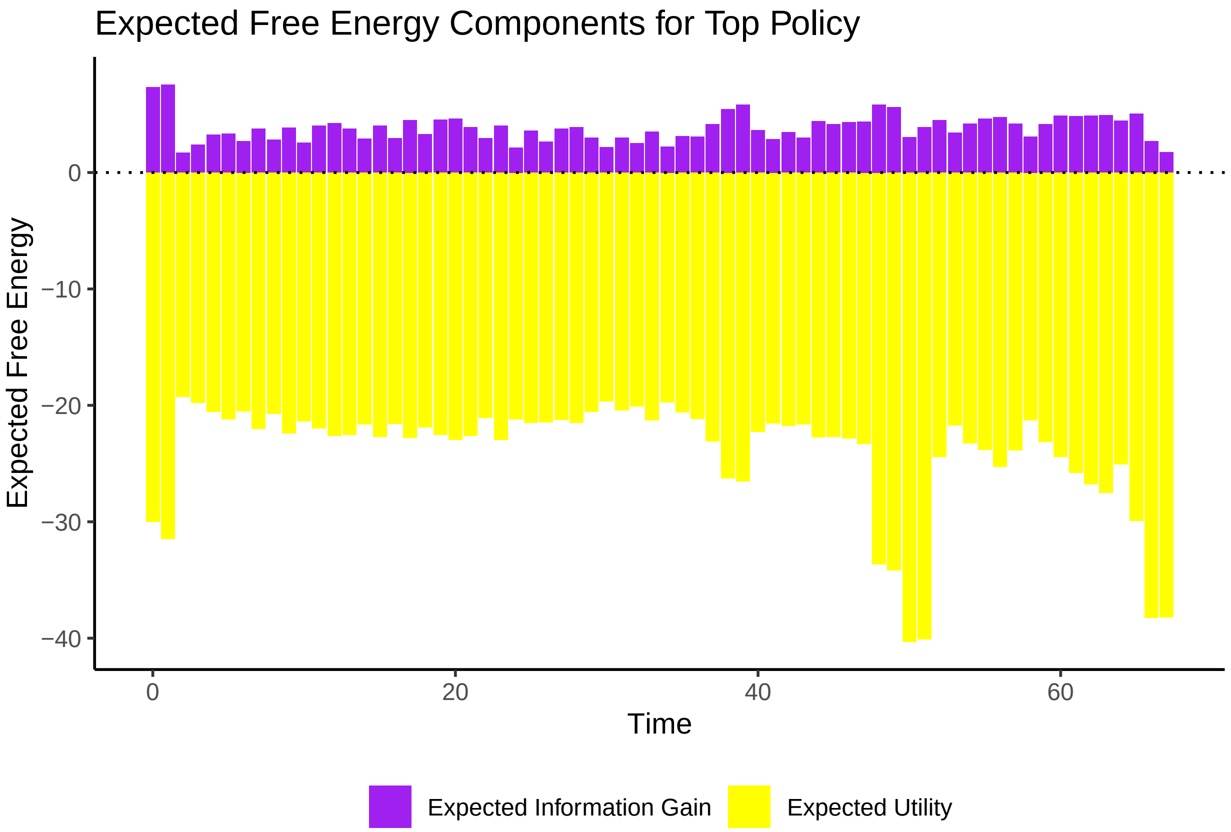
Policy evalutation



The above plot shows the highest evaluated policy at timestep 16 and at timestep 34. At timestep 16 not treatment is planned, and testing is withdrawn for next timestep but picked up again for reaminder of the period. For 34 treatment is suggested applied for three rounds. Note that policies are evaluated each timestep, and these plans are likely not finished. Instead a course of action is decided on each turn. The highest rated policy for each timestep is plotted



The above plot doesnt show realized actions but cumultatiove best action from perspective of each time point. This means that for completely blank spots, there is no point where model thinks treating or testing would be the optimal course of action. This doesn't never happen for treating, but some periods of treatment are throught time best held withdrawn. A darker shade indicate that on multiple prior timesteps, the model thought acting would be best course of action. Excact decision making for each policy can also be extracted. Since each policy is evaluated for its EFE, its expected information and utility gains can be extracted

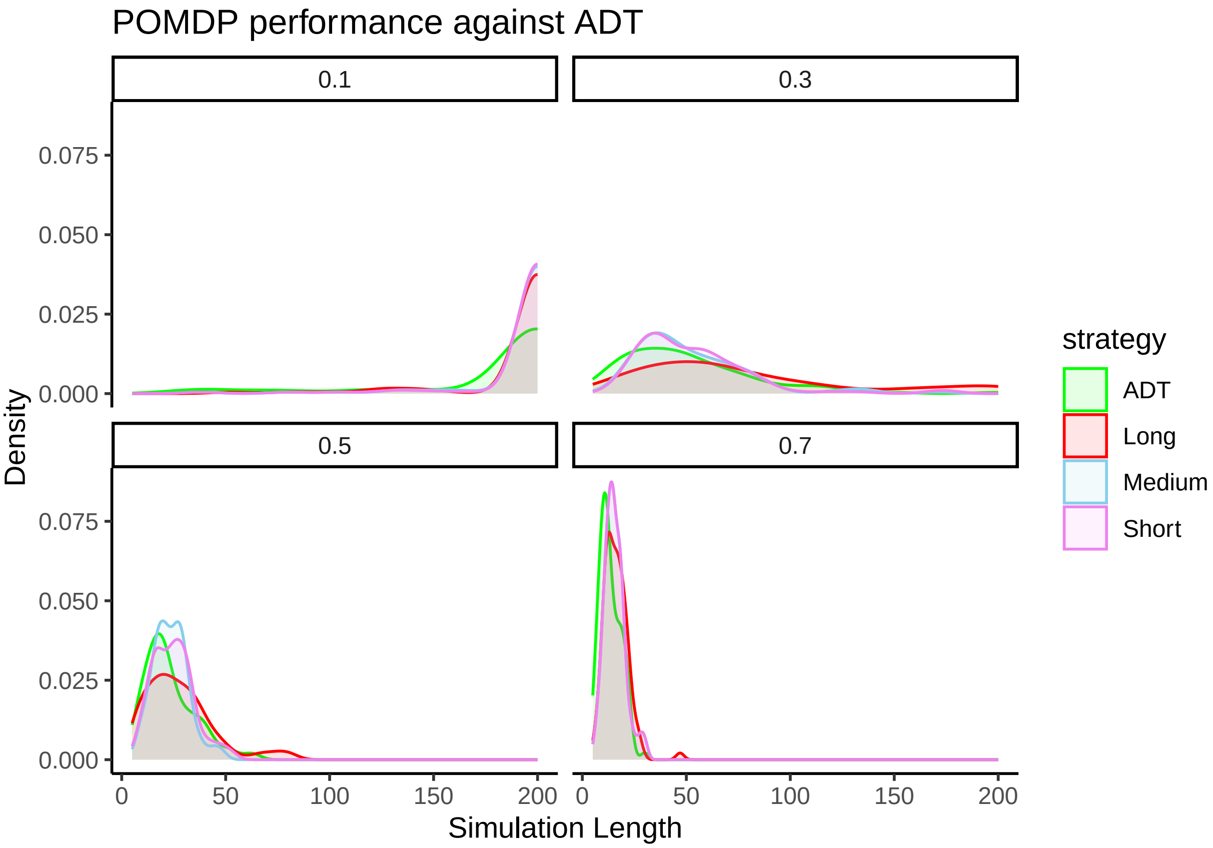


Performance simulation, where a range-bounded approach to managing the tumor is used, where the treatment is begun each time the tumor state increases above level 3, and is withdrawn when the tumor state drops below 2. 100 runs at 4 different probabilities of increasing the tumor state is run for maximally 200 timesteps. For each run, vectors of outcomes are pregenerated, meaning that vector of whether the tumor increases of length 200 is predetermined where each entry has the probability specified for the entire simulation. Likewise vectors are generated for treatment outcomes at each resistance level, and outcomes for resistance drops. This is done to ensure comparibility between the range-bounded approach and POMDP models that consider different future outcomes at different lengths. In these runs the tumor state is perfectly observable, meaning a tumor state always generates tumor observation that corresponds to hidden factor. The POMPD can only observe the tumor level, and its prior preferences are uniform over all tumor states except for the highest. These POMPDS only consider a combination of policies at each timestep. They can only consider treating or not treating for three timesteps in a row. The shortest horizon model only considers one of these blocks, while the medium considers two blocks of treating or not treating, for total horizon of 6 steps into the future. The longest horizon model considers three blocks for a total of 9 steps into the future. This is done to ease the computational burden, since it greatly reduces the total number of policies to be evaluated.

Results

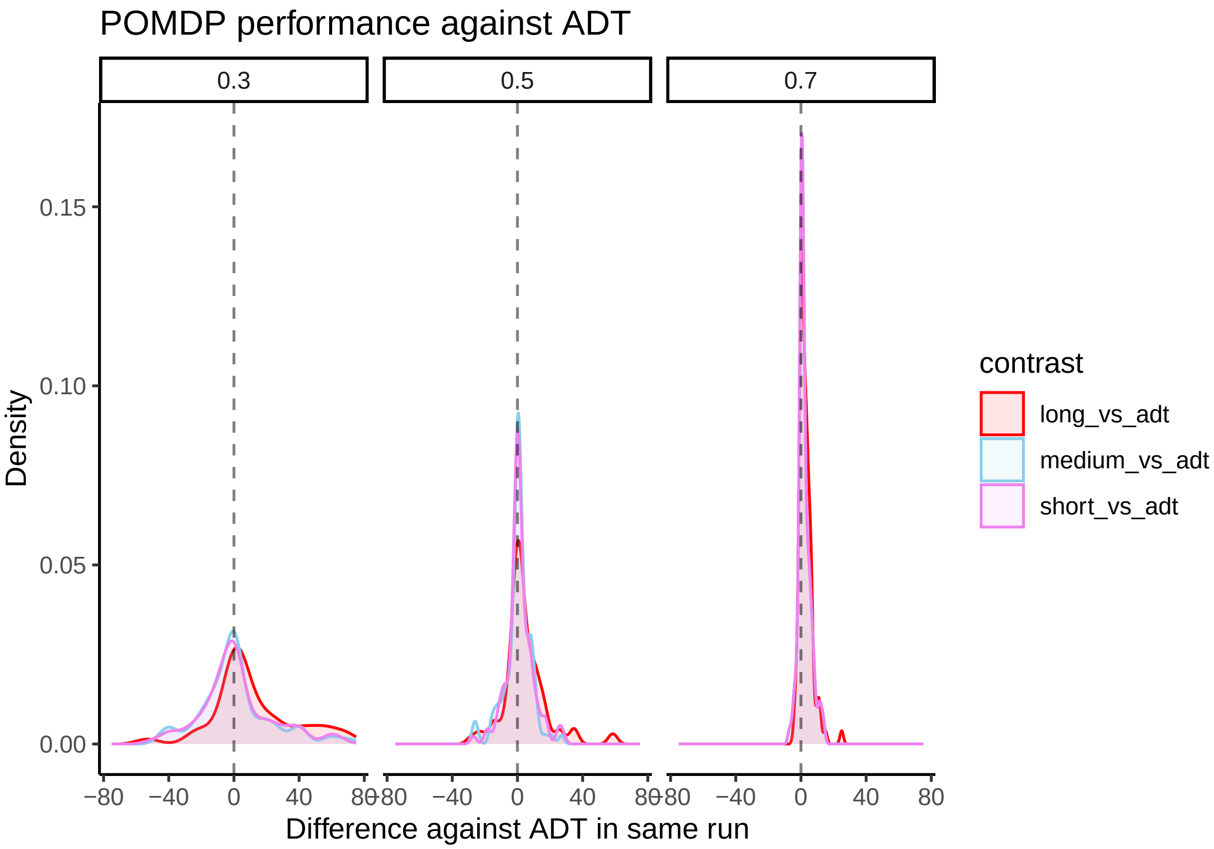
Capabilities

Performance of POMDPs vs ADT heuristic

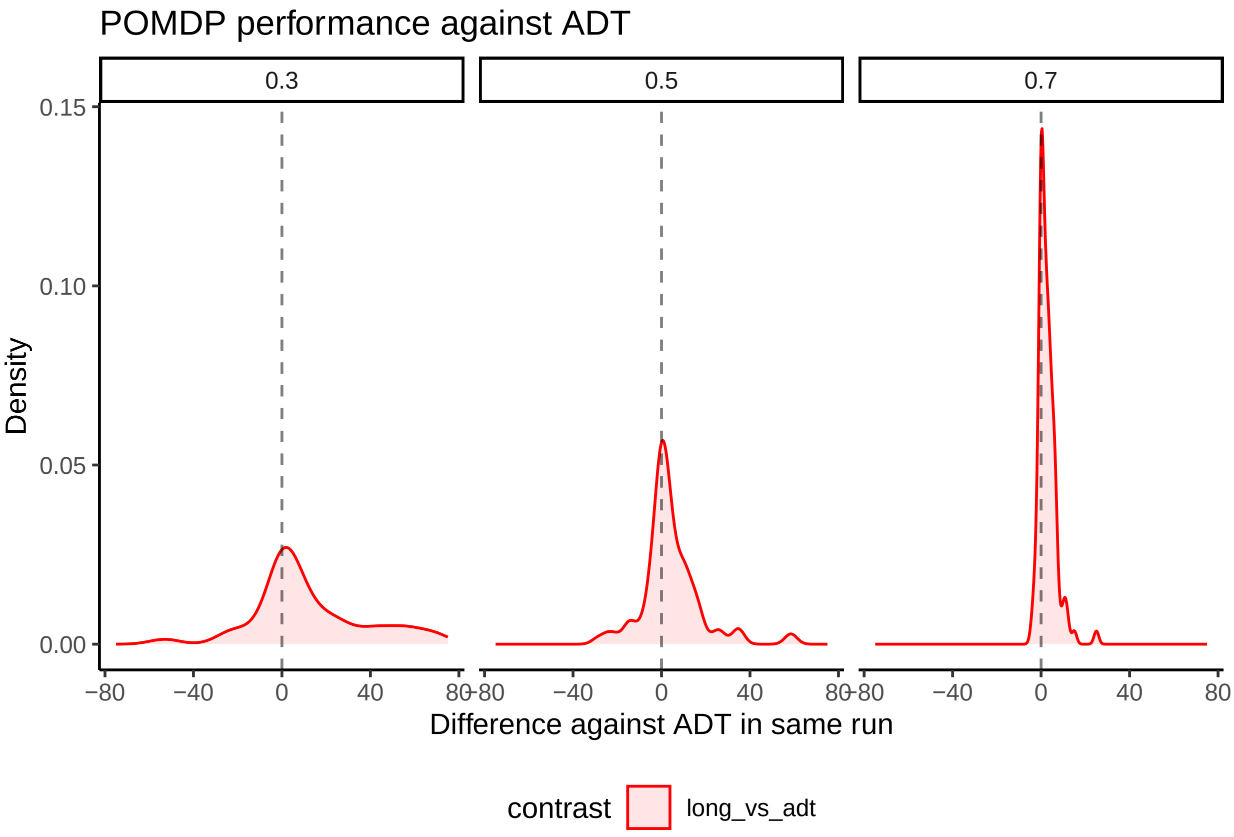


Four different simulation runs each consisting of 100 runs, of maximally 200 timesteps for different rates of tumor growth (respectively 1. .3, .5, and .7 probability of increasing the tumor state) . On each run the the outcomes on each of the potential 200 random variables are generated and each strategy is therefore tested on a the same enviroment. On the lowest tumor growth rate almost all the runs reach maximal length.

Contrasts against ADT-performance for growth rate .3, .5 and .7 are plotted. .1 is omitted since it appears that maximal perforamnce was achieved for each policy horizon of the POMPD strategies.



Two interpretations emerge from the contrasted simulation lengths. As the the growth rate increases, the POMDP seems to run for longer than the specific instance of the range bounded approach. It also seems that longer policy horizon seems beneficial. When the tumor state had .7 probability of increasing each timestep, the longest POMPD with the longest policy horizon performed as follows.



At more aggressive cancer rates, the simulation consistently to outperform the range bounded approach. The absoulute increase in timesteps decreases however.

Discussion

What has this project found

The present project tested a simple modification to how POMDPs typically are implemented in Active Inference. This was done to investigate whether POMDPs could learn an control a hidden state that could only be inferred on the basis of its consequences on other states. This was done in a simulated environment designed to mimic the dynamics of adaptive therapy since resistance dynamics of cancers are not directly orbservable, but have to be inferred through treatment response. The modification was successfull, allowing the POMDP to model an underlying resistance state that controlled the efficacy of treatment, despite the underlying state not producing any observable signals itself. POMPDs were further tested against a range-bounded treatment strategy, and were found to outperfom the range-bounded strategy. The difference in performance increased as the aggresivity of tumor growth increased and the future time-horizon that the POMPDs considered at each time step. The finding underscores the prospective of using the paradigm of Active Inference in general to plan adaptive therapy, and in particular has shown that under simplifed circumstances, POMDPs could be a useful choice of model to implement Active Inference for real-time treatment decision making.

Additionally, the paper demonstrates that active inference implementation of POMPS show a number of desirable qualities for adaptive therapy POMDPs:

- The models are flexible. Computation and our ability to inform likelihood and transition dynamics, are the only limit to the combinations of hidden states, actions and outcome modalities. This present paper implemented a pomp capapble of combining multiple decisions, treating and testing, but more complex generative models could including multiple testing and treatment actions. Likewise, as (West et al. 2023) highlight desiderata of biological factors that would be important to incorporate into modelling decisions.

- A second quality is the capacity to balance expected information gain is balanced against expected utility. The certainty of the model is incorpareted directly in selecting when to apply treatments and when to test. From these considerations suggest that there is more information to be gained from testing when treatment is being applied, since it provides a view into the underlying resistance dynamics. Immediate testing seem to suggest that the models are more likely to apply testing when also applying treatment. Another interesting corrolary is that a POMPD could believe that two patients have exactly the same tumor burden but suggest treatment for one, but and not the other. This a feature, not a bug. If the model is certain about an underlying state (e.g. a resistance for the models in the present paper) there is less gain in information from treating. On the other hand, if the model is uncertain about the resistance state, it might be worthwhile to treat, simply for gain in information. This is appropriate behavior, since it allows the model better is chances of keeping the patient alive in the future.
- Another quality of models desired (West et al. 2023) is the rigorous uncertainty. Since the POMDP's certainty about current and future states is easily extractable, this criteria is arguably met. Additionally, its "reasoning", i.e. free energy estimates can be examined for each action.

Future research

While the results are initially promising, the simple nature of the simulation prohibits drawing any strong conclusions about whether active inference POMPDs eventually could come to be a good model choice in a clinical setting.

To further investigate, whether POMDPs could apply to a clinical setting multiple problems will have to be solved. For example in the present study the environment featured only discrete values. Since bio-markers and dosage intensities likely will be continuous values, some combination of binning continuous values to discrete values or adapting he model structure to work with continuous data is necessary. Deep-learning have for example been used to construct likelihood mappings and transition probabilities in POMDPs Çatal et al. (2020) from the continous data. Another key issue, is determining a way to inform likelihood and transition probablities in a fashion that would be viable in a clinical setting. The present study, simply used the actual transition probabilities of the simulated enviroment, and depeding on the simulation run also used un-noised likelihood mappings. Possibly, simulations of cancer dynamics, such as (Zhang et al. 2017) could be used to inform transisition probabilities, and a clinical model could potentially readjust to patient data. The Active Inference implementation of POMDPs have a developed litterature on learning probabilities from data and even how to incorporate the information gain of learning transition and likelihood probabilities into decision-making ((Smith, Friston, and Whyte 2022) and Da Costa et al. (2020)]). Another key issue, is constructing a more rigourous benchmarking system for proposed models. In the present paper, the ranges of the compared range-bounded model, were selected fairly arbitrariliy. It produced convincing results during intial testing, and did manage to control tumor levels for substantial amount of time. However, a systematic method of comparing models is necessary, espicially considering the difficulty of real-world tests. No matter the choice of modelling framework, building a bench-marking suite must be crucial for the adaptive therapy disicpline. Considering that difficulty of real-world testing, we will have to maximize the infomration that can be extracted from simulation work. (West et al. 2023) have produced an a detailed qualitative account of needed developments in mathematical models. Translating this account to a set of simulation environments would be extremely usefull. A number of usefull simulation-script propably already exists. Zhang et al. (2017) for example released their simulations from matlab. Wrapping already existing simulations into enviroments that can easily exchange actions and orbsevations. The API used in Gymnasium (formerly OpenAI Gym) (Towers et al. 2024) could for example be used to minimize the friction for non-oncology researchers. If experts in adaptive therapy were to predetermine a set of bench marks, it would also greatly ease the burden of outsiders to first identify what even would be a usefull contribution and instead free more time for them to implement it.

If an easy to use bench-marking suite existed, POMPDs could also be "reverse-engineered" from more complex models. If a black-box model, like a neural network can be shows can be shown to succesuffly control treatment application in more complicated simulations of cancer dynamics. This would presumably not be too difficult considering that POMPDs have often been fitted to human decision-making in computational psychiatry. The same techniques could potentially allows us translate the decision making of a black-box model into the structure of a POMPD, thus combining the performance of the black-box model with the transparency of POMPD structure.

Åström, K. J. 1969. "Optimal Control of Markov Processes with Incomplete State-Information II. The Convexity of the Lossfunction." *Journal of Mathematical Analysis and Applications* 26 (2): 403–6. https://doi.org/10.1016/0022-247X(69)90163-2.

Bukowski, Karol, Mateusz Kciuk, and Renata Kontek. 2020. "Mechanisms of Multidrug Resistance in Cancer Chemotherapy." *International Journal of Molecular Sciences* 21 (9): 3233. https://doi.org/10.3390/ijms21093233.

Çatal, Ozan, Samuel Wauthier, Cedric De Boom, Tim Verbelen, and Bart Dhoedt. 2020. "Learning Generative State Space Models for Active Inference." *Frontiers in Computational Neuroscience* 14 (November). https://doi.org/10.3389/fncom.2020.574372.

Center, H. Lee Moffitt Cancer, and Research Institute. 2024. "A Pilot Study of Adaptive Abiraterone Therapy for Metastatic Castration Resistant Prostate Cancer." https://clinicaltrials.gov/study/NCT02415621.

Da Costa, Lancelot, Thomas Parr, Noor Sajid, Sebastijan Veselic, Victorita Neacsu, and Karl Friston. 2020. "Active Inference on Discrete State-Spaces: A Synthesis." *Journal of Mathematical Psychology* 99 (December): 102447. https://doi.org/10.1016/j.jmp.2020.102447.

Friston, Karl J., Jean Daunizeau, and Stefan J. Kiebel. 2009. "Reinforcement Learning or Active Inference?" *PLoS ONE* 4 (7): e6421. https://doi.org/10.1371/journal.pone.0006421.

Gatenby, Robert A., Ariosto S. Silva, Robert J. Gillies, and B. Roy Frieden. 2009. "Adaptive Therapy." *Cancer Research* 69 (11): 4894–4903. https://doi.org/10.1158/0008-5472.CAN-08-3658.

Hansen, Elsa, and Andrew F. Read. 2020. "Modifying Adaptive Therapy to Enhance Competitive Suppression." *Cancers* 12 (12): 3556. https://doi.org/10.3390/cancers12123556.

Heins, Conor, Beren Millidge, Daphne Demekas, Brennan Klein, Karl Friston, Iain D. Couzin, and Alexander Tschantz. 2022a. "Pymdp: A Python Library for Active Inference Indiscrete State Spaces." *Journal of Open Source Software* 7 (73): 4098. https://doi.org/10.21105/joss.04098.

Heins, Conor, Beren Millidge, Daphne Demekas, Brennan Klein, Karl Friston, Iain Couzin, and Alexander Tschantz. 2022b. "Pymdp: A Python Library for Active Inference in Discrete State Spaces." *Journal of Open Source Software* 7 (73): 4098. https://doi.org/10.21105/joss.04098.

Kaelbling, Leslie Pack, Michael L. Littman, and Anthony R. Cassandra. 1998. "Planning and Acting in Partially Observable Stochastic Domains." *Artificial Intelligence* 101 (1): 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X.

Smith, Ryan, Karl J. Friston, and Christopher J. Whyte. 2022. "A Step-by-Step Tutorial on Active Inference and Its Application to Empirical Data." *Journal of Mathematical Psychology* 107 (April): 102632. https://doi.org/10.1016/j.jmp.2021.102632.

Staňková, Kateřina, Joel S. Brown, William S. Dalton, and Robert A. Gatenby. 2019. "Optimizing Cancer Treatment Using Game Theory." *JAMA Oncology* 5 (1): 96–103. https://doi.org/10.1001/jamaoncol.2018.3395.

Towers, Mark, Jordan K Terry, Ariel Kwiatkowski, John U. Balis, Gianluca Cola, Tristan Deleu, Manuel Goulão, et al. 2024. *Gymnasium*. Zenodo. https://doi.org/10.5281/zenodo.10655021.

West, Jeffrey, Fred Adler, Jill Gallaher, Maximilian Strobl, Renee Brady-Nicholls, Joel Brown, Mark Roberson-Tessi, et al. 2023. "A Survey of Open Questions in Adaptive Therapy: Bridging Mathematics and Clinical Translation." Edited by Richard M White. *eLife* 12 (March): e84263. https://doi.org/10.7554/eLife.84263.

Zhang, Jingsong, Jessica J. Cunningham, Joel S. Brown, and Robert A. Gatenby. 2017. "Integrating Evolutionary Dynamics into Treatment of Metastatic Castrate-Resistant Prostate Cancer." *Nature Communications* 8 (1): 1816. https://doi.org/10.1038/s41467-017-01968-5.