Putting it all Together

Introduction to Automated Science

SLAS 2024

Batched experiments

In theory, performing one experiment at a time maximizes the information gained. These experiments leverage all existing knowledge during planning.

Batched experiments

In theory, performing one experiment at a time maximizes the information gained. These experiments leverage all existing knowledge during planning.

In practice, a sequence of individual runs is wildly inefficient.

- Experiments require long (e.g. overnight) incubation.
- Single experiments do not leverage automation or high-throughput instruments.
- Iterating between modeling and experiments creates downtime.

Batched experiments

In theory, performing one experiment at a time maximizes the information gained. These experiments leverage all existing knowledge during planning.

In practice, a sequence of individual runs is wildly inefficient.

- Experiments require long (e.g. overnight) incubation.
- Single experiments do not leverage automation or high-throughput instruments.
- Iterating between modeling and experiments creates downtime.

Planning **batch** or **parallel** experiments is challenging since most planning algorithms find the best treatment independent of other runs.

Batched experiments (continued)

To avoid redundant information, batch planning combines two objectives:

1. Maximize a **planning metric** (response = exploitation, uncertainty = exploration) for every point in the batch.
2. Maximize the **diversity** of the points in the batch (e.g. the minimum distance between any treatments).

Batching reduces the information gained *per run*; however, batched experiments are less expensive and may increase the overall run budget.

Remember the goal is to maximize **total information**:

$$\text{total information} = (\text{information per run}) \times (\text{\# of runs}).$$

Automation challenges

To maximize the information gained, the planning agent should be allowed to select **any treatment** during each round of experiments.

Automation challenges

To maximize the information gained, the planning agent should be allowed to select **any treatment** during each round of experiments.

This requires experiments “on demand” instead of a structured high-throughput screen.

Automation challenges

To maximize the information gained, the planning agent should be allowed to select **any treatment** during each round of experiments.

This requires experiments "on demand" instead of a structured high-throughput screen.

For example, an agent that can only test *pairs* of small molecules can only learn *two-way interactions* among the compounds.

Automation challenges

To maximize the information gained, the planning agent should be allowed to select **any treatment** during each round of experiments.

This requires experiments “on demand” instead of a structured high-throughput screen.

For example, an agent that can only test *pairs* of small molecules can only learn *two-way interactions* among the compounds.

In practice, continuous variables cannot be varied freely (dilutions of a compound), and not every factor can be varied for every run (temperature in the wells of a multiwell plate).

Automation challenges

To maximize the information gained, the planning agent should be allowed to select **any treatment** during each round of experiments.

This requires experiments “on demand” instead of a structured high-throughput screen.

For example, an agent that can only test *pairs* of small molecules can only learn *two-way interactions* among the compounds.

In practice, continuous variables cannot be varied freely (dilutions of a compound), and not every factor can be varied for every run (temperature in the wells of a multiwell plate).

You need to push your automation team to increase flexibility **even at the expense of throughput**.

The Automated Science checklist

1. Determine the **response** and **factors**. Are you optimizing or characterizing the response?

The Automated Science checklist

1. Determine the **response** and **factors**. Are you optimizing or characterizing the response?
2. Meet with the automation team and decide
   - the **range** and **levels** for each factor
   - treatment **flexibility** (how independently can you vary each factor)
   - **run budget** = initial runs + (rounds $\times$ runs/round)

The Automated Science checklist

1. Determine the **response** and **factors**. Are you optimizing or characterizing the response?
2. Meet with the automation team and decide
    - the **range** and **levels** for each factor
    - treatment **flexibility** (how independently can you vary each factor)
    - **run budget** = initial runs + (rounds $\times$ runs/round)
3. Meet with the modeling team and decide
    - type of **model** and method for **uncertainty** quantification
    - planning **policy** (continuous vs. discrete, mean response vs. EI)

The Automated Science checklist

1. Determine the **response** and **factors**. Are you optimizing or characterizing the response?
2. Meet with the automation team and decide
   - the **range** and **levels** for each factor
   - treatment **flexibility** (how independently can you vary each factor)
   - **run budget** = initial runs + (rounds $\times$ runs/round)
3. Meet with the modeling team and decide
   - type of **model** and method for **uncertainty** quantification
   - planning **policy** (continuous vs. discrete, mean response vs. EI)
4. Meet with both the automation and modeling teams to decide
   - initial runs—here you can limit flexibility to increase throughput or lower cost
   - QC strategy
   - turnaround times (plans $\rightarrow$ data; data $\rightarrow$ plans)

The Automated Science checklist

1. Determine the **response** and **factors**. Are you optimizing or characterizing the response?
2. Meet with the automation team and decide
   - the **range** and **levels** for each factor
   - treatment **flexibility** (how independently can you vary each factor)
   - **run budget** = initial runs + (rounds $\times$ runs/round)
3. Meet with the modeling team and decide
   - type of **model** and method for **uncertainty** quantification
   - planning **policy** (continuous vs. discrete, mean response vs. EI)
4. Meet with both the automation and modeling teams to decide
   - initial runs—here you can limit flexibility to increase throughput or lower cost
   - QC strategy
   - turnaround times (plans $\rightarrow$ data; data $\rightarrow$ plans)
5. Automate!