UFZ HELMHOLTZ Zentrum für Umweltforschung

Automated gating of flow cytometry data

Carmen Bruckmann

tbi Winterseminar 2023

Leipzig, February 16, 2023

Introduction to flow cytometry

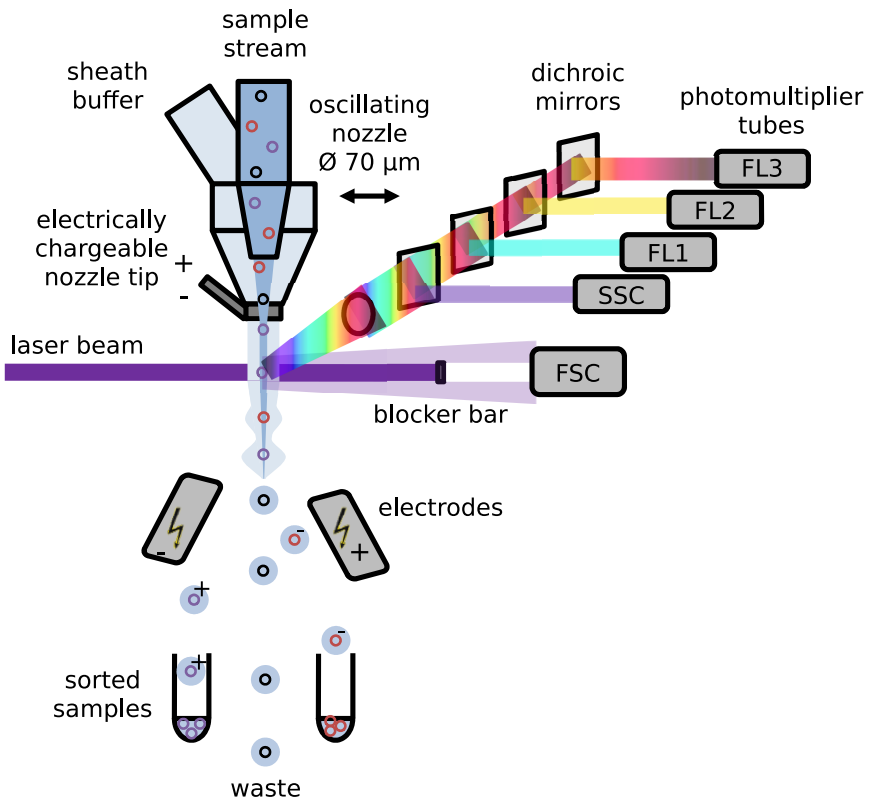


Figure: Adapted from: J. Lambrecht, **The community sensor - Monitoring and control of microbiome dynamics in anaerobic processes**, *PhD Dissertation, 2020*

Introduction to flow cytometry

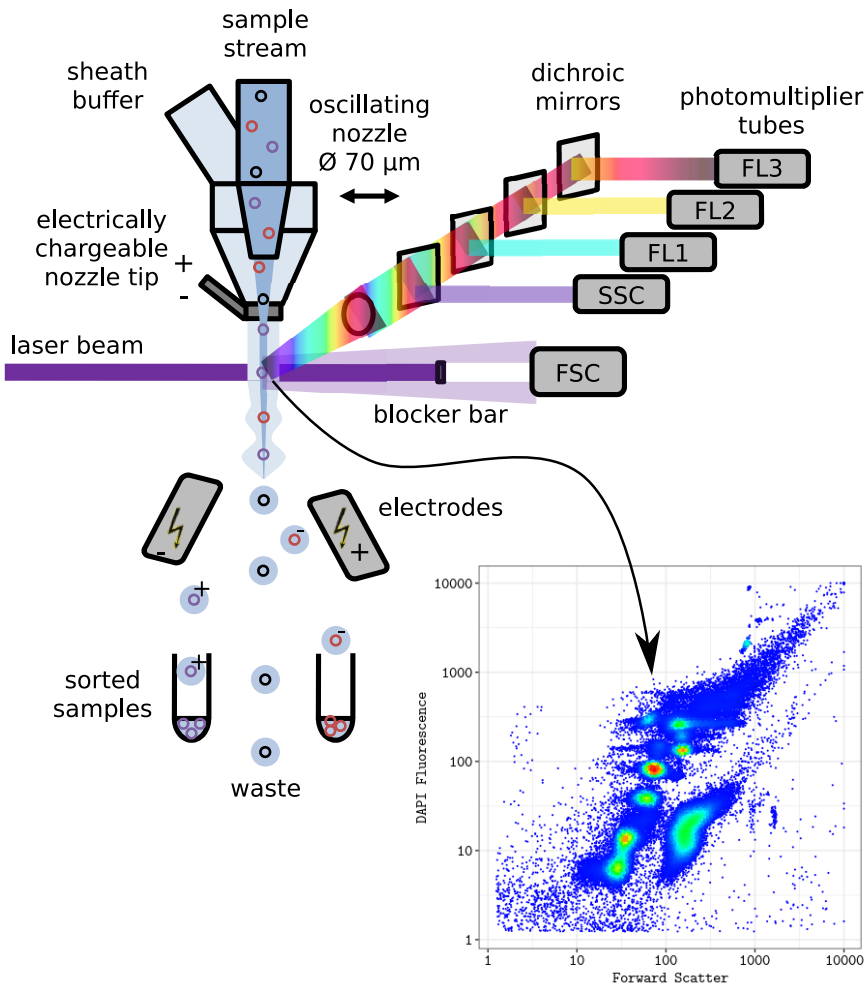
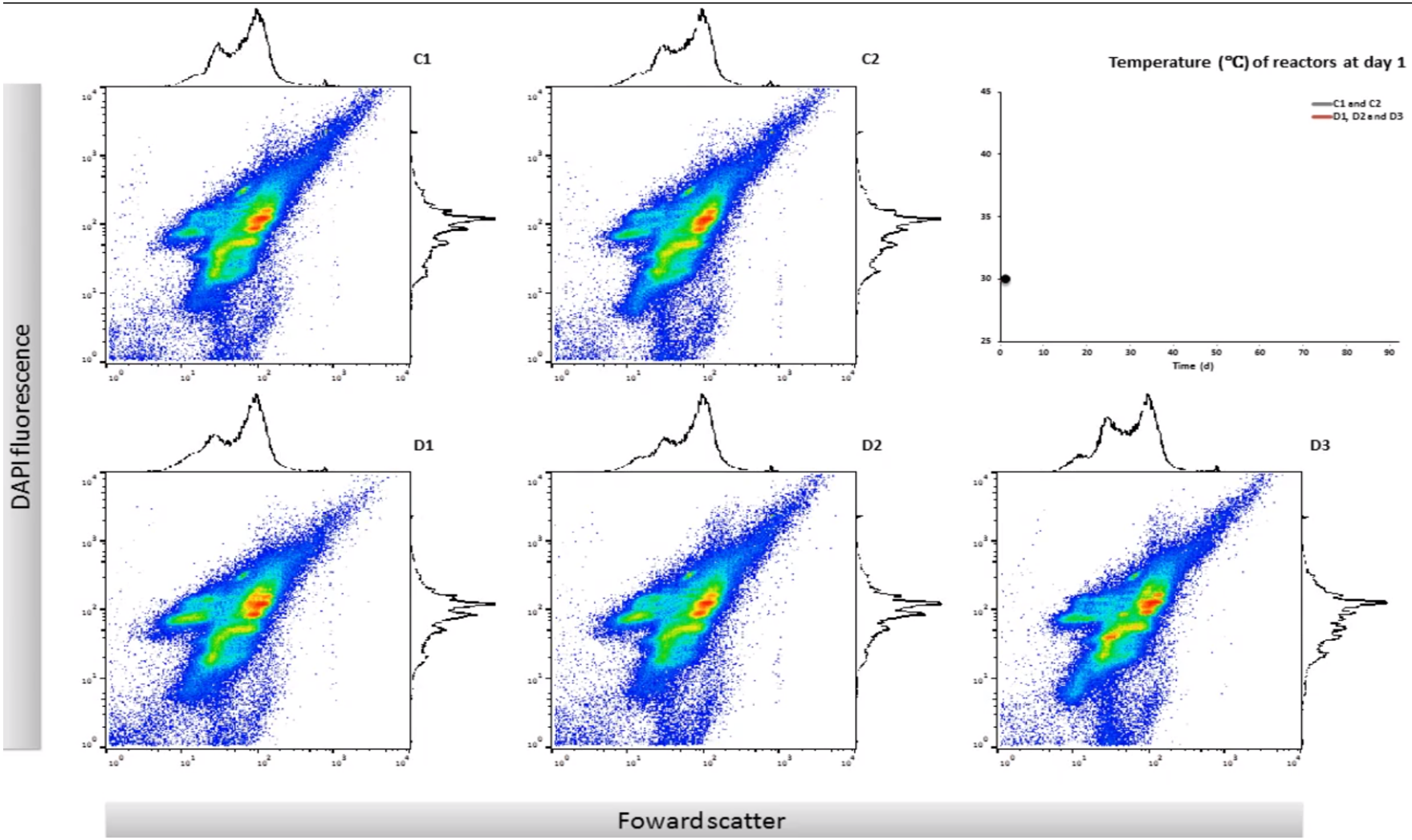


Figure: Adapted from: J. Lambrecht, **The community sensor - Monitoring and control of microbiome dynamics in anaerobic processes**, *PhD Dissertation, 2020*

Introduction to flow cytometry



Liu et al., **Neutral mechanisms and niche differentiation in steady-state insular microbial communities revealed by single cell analysis**, *Environmental Microbiology, 2019*

Manual Gating

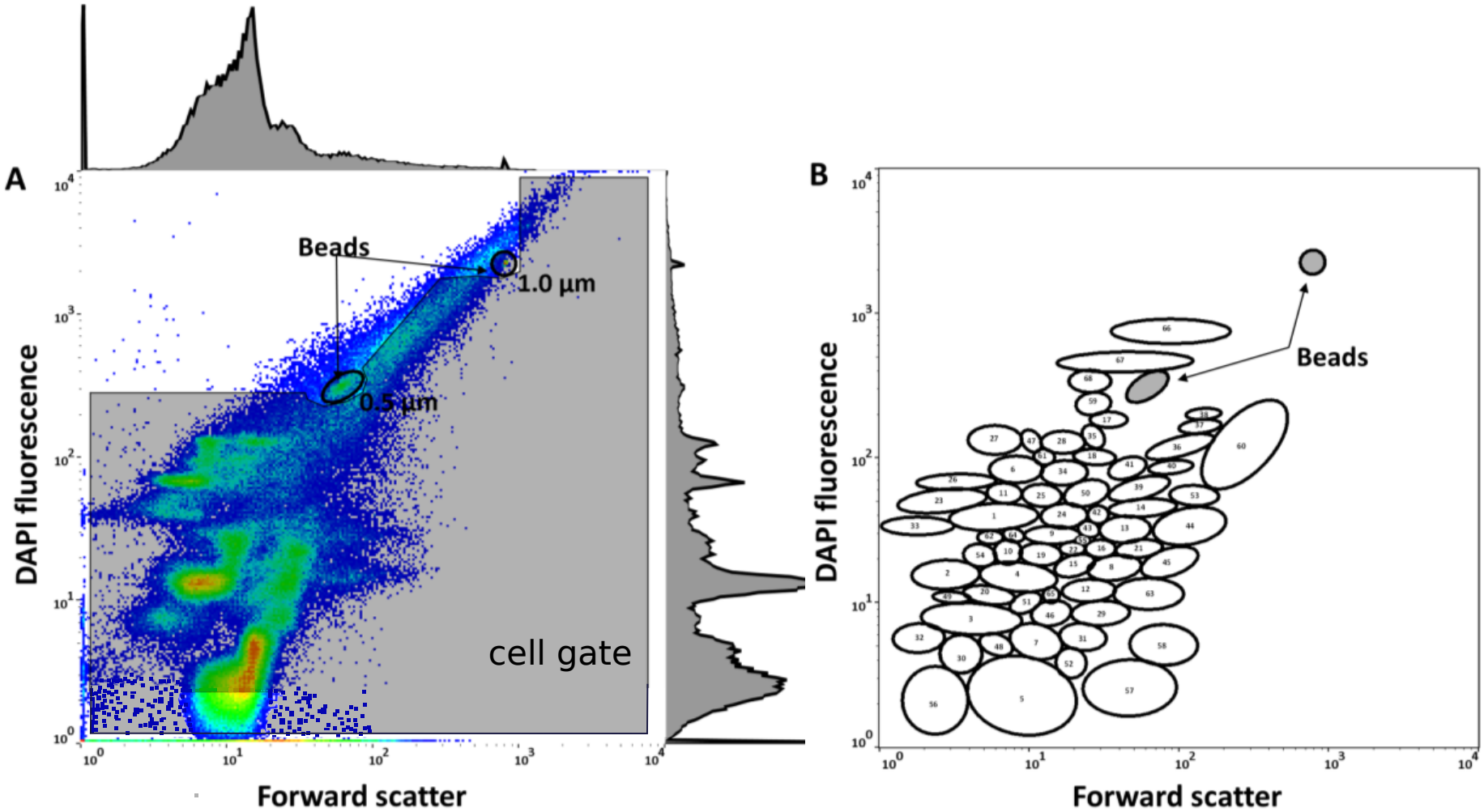


Figure: Liu et al., **Neutral mechanisms and niche differentiation in steady-state insular microbial communities revealed by single cell analysis**, *Environmental Microbiology, 2019*

Underlying ecological questions

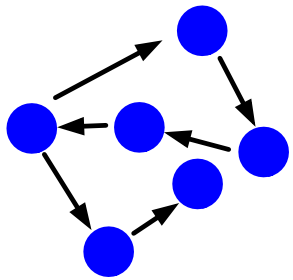
- How diverse is the community?

Underlying ecological questions

- How diverse is the community?
- How does the microbial community develop over time?

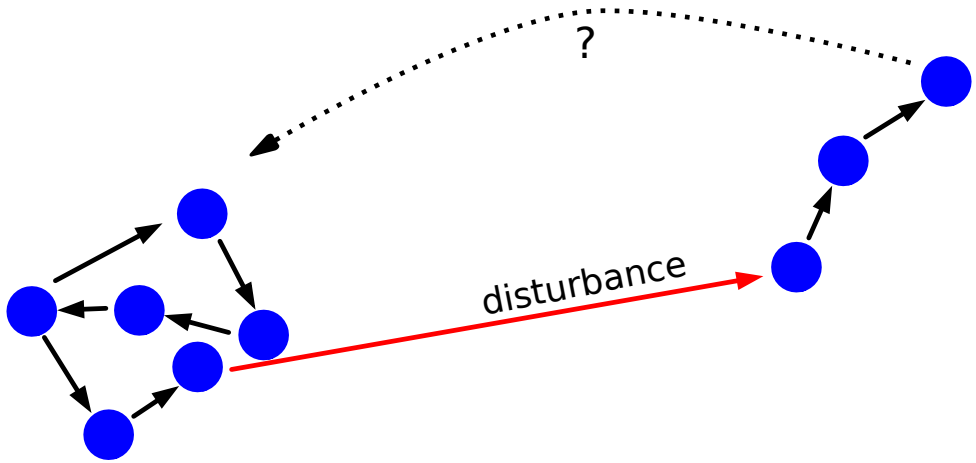
Underlying ecological questions

- How diverse is the community?
- How does the microbial community develop over time?



Underlying ecological questions

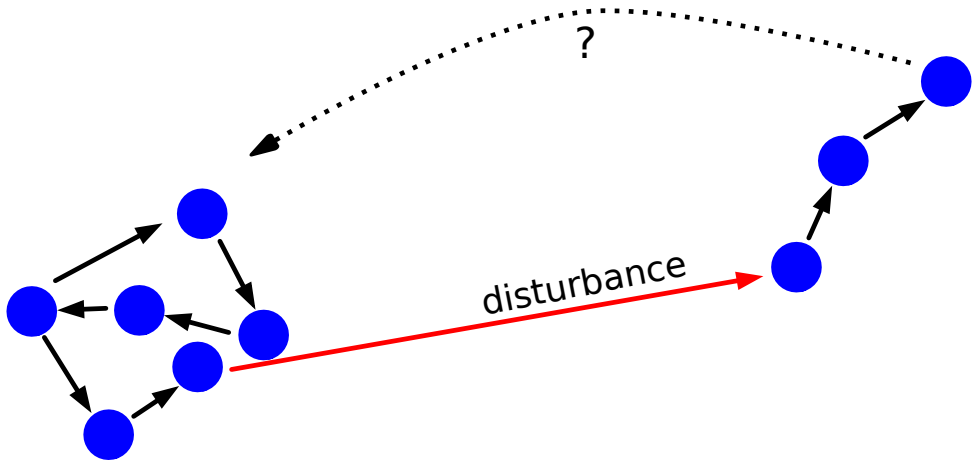
- How diverse is the community?
- How does the microbial community develop over time?
- Is the community able to recover from a disturbance?
- ...



Underlying ecological questions

- How diverse is the community?
- How does the microbial community develop over time?
- Is the community able to recover from a disturbance?
- ...

For all of this we need the cell numbers per subcommunity over time, and thus **gating**



Underlying ecological questions

- How diverse is the community?
- How does the microbial community develop over time?
- Is the community able to recover from a disturbance?
- ...

For all of this we need the cell numbers per subcommunity over time, and thus **gating**

Gating is still done by hand but this is

- user dependent

Underlying ecological questions

- How diverse is the community?
- How does the microbial community develop over time?
- Is the community able to recover from a disturbance?
- ...

For all of this we need the cell numbers per subcommunity over time, and thus **gating**

Gating is still done by hand but this is

- user dependent
- not reproducible

Underlying ecological questions

- How diverse is the community?
- How does the microbial community develop over time?
- Is the community able to recover from a disturbance?
- ...

For all of this we need the cell numbers per subcommunity over time, and thus **gating**

Gating is still done by hand but this is

- user dependent
- not reproducible
- time-consuming

Underlying ecological questions

- How diverse is the community?
- How does the microbial community develop over time?
- Is the community able to recover from a disturbance?
- ...

For all of this we need the cell numbers per subcommunity over time, and thus **gating**

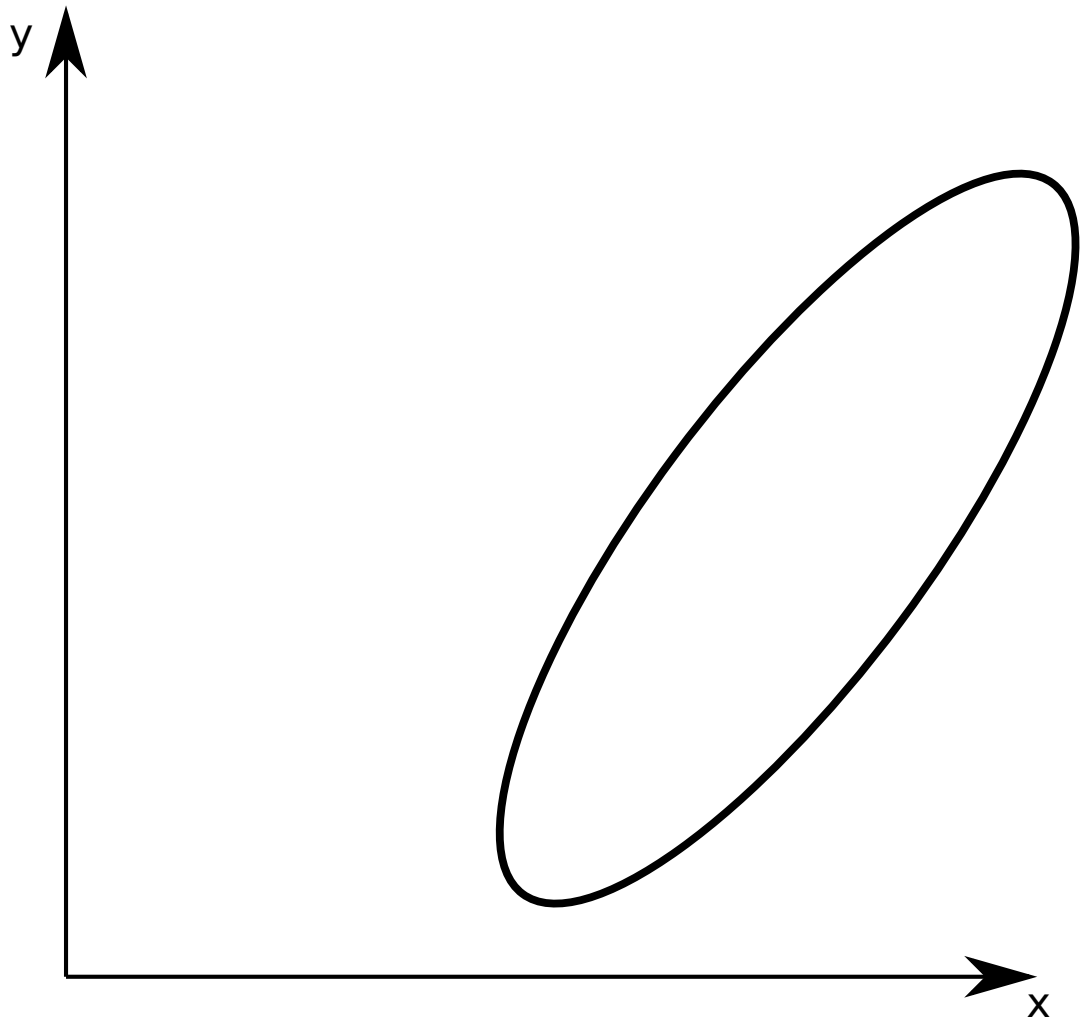
Gating is still done by hand but this is

- user dependent
- not reproducible
- time-consuming

$\Rightarrow$ **Need of automated gating!**

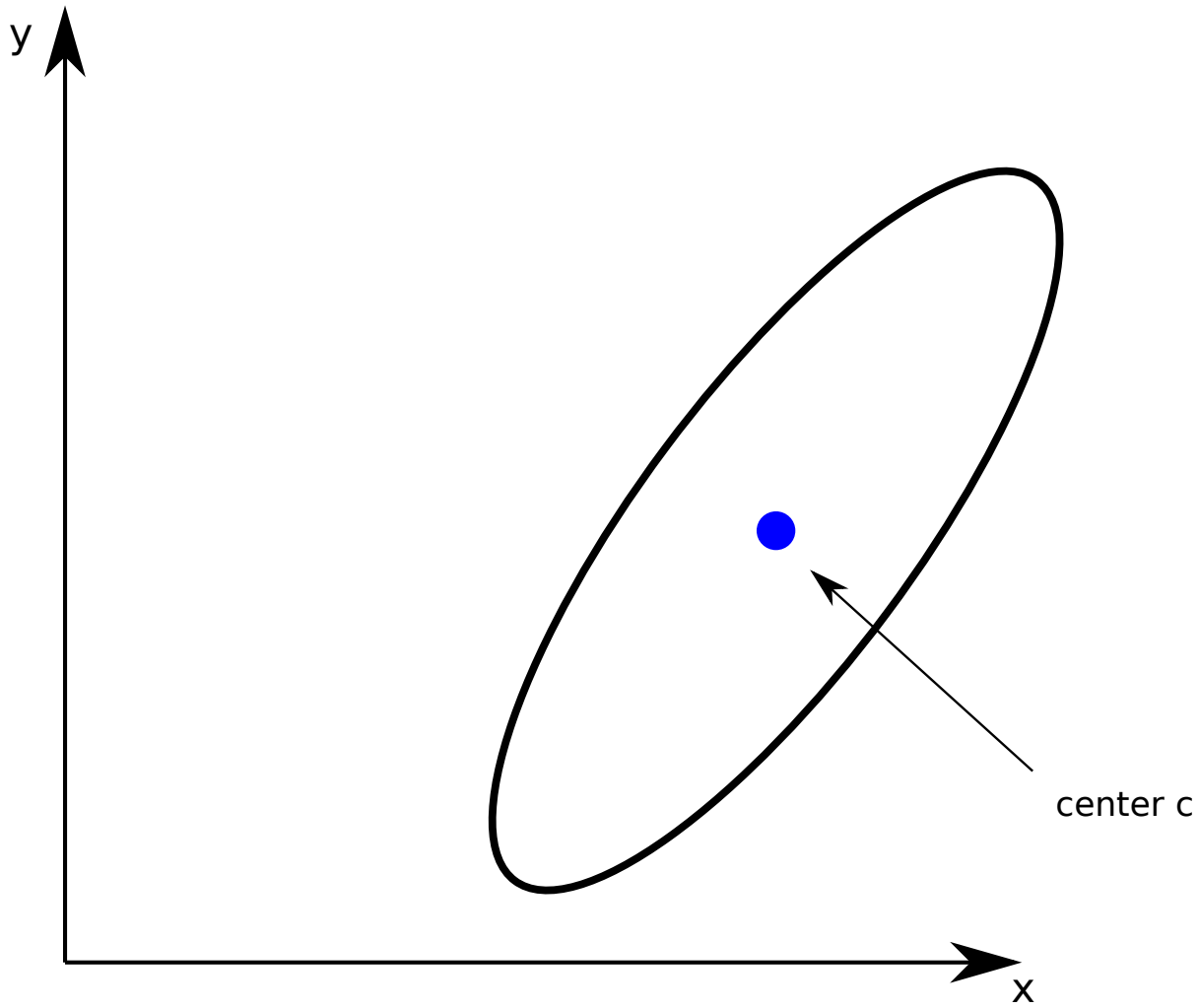
Introduction

What is an ellipse?



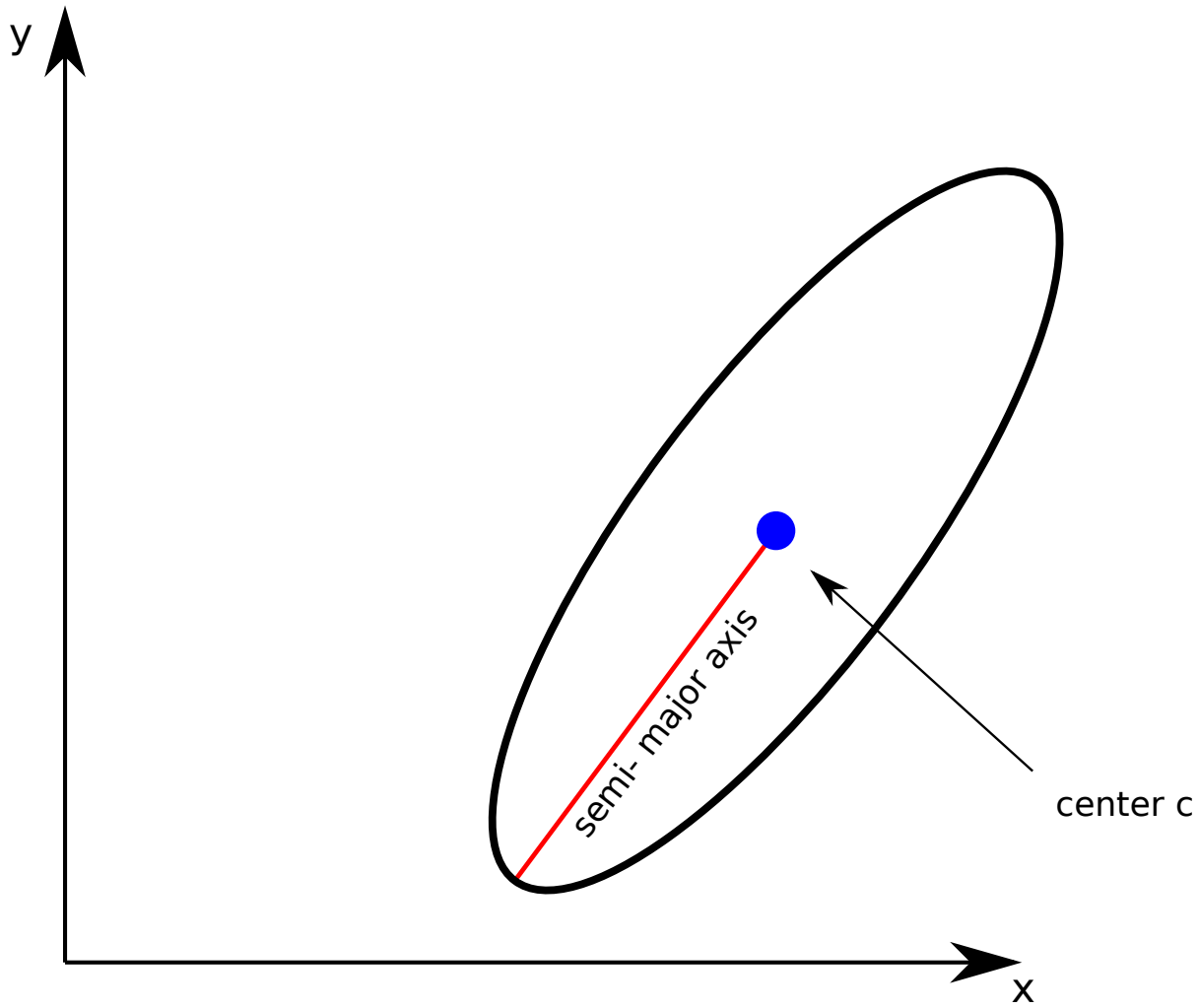
Introduction

What is an ellipse?



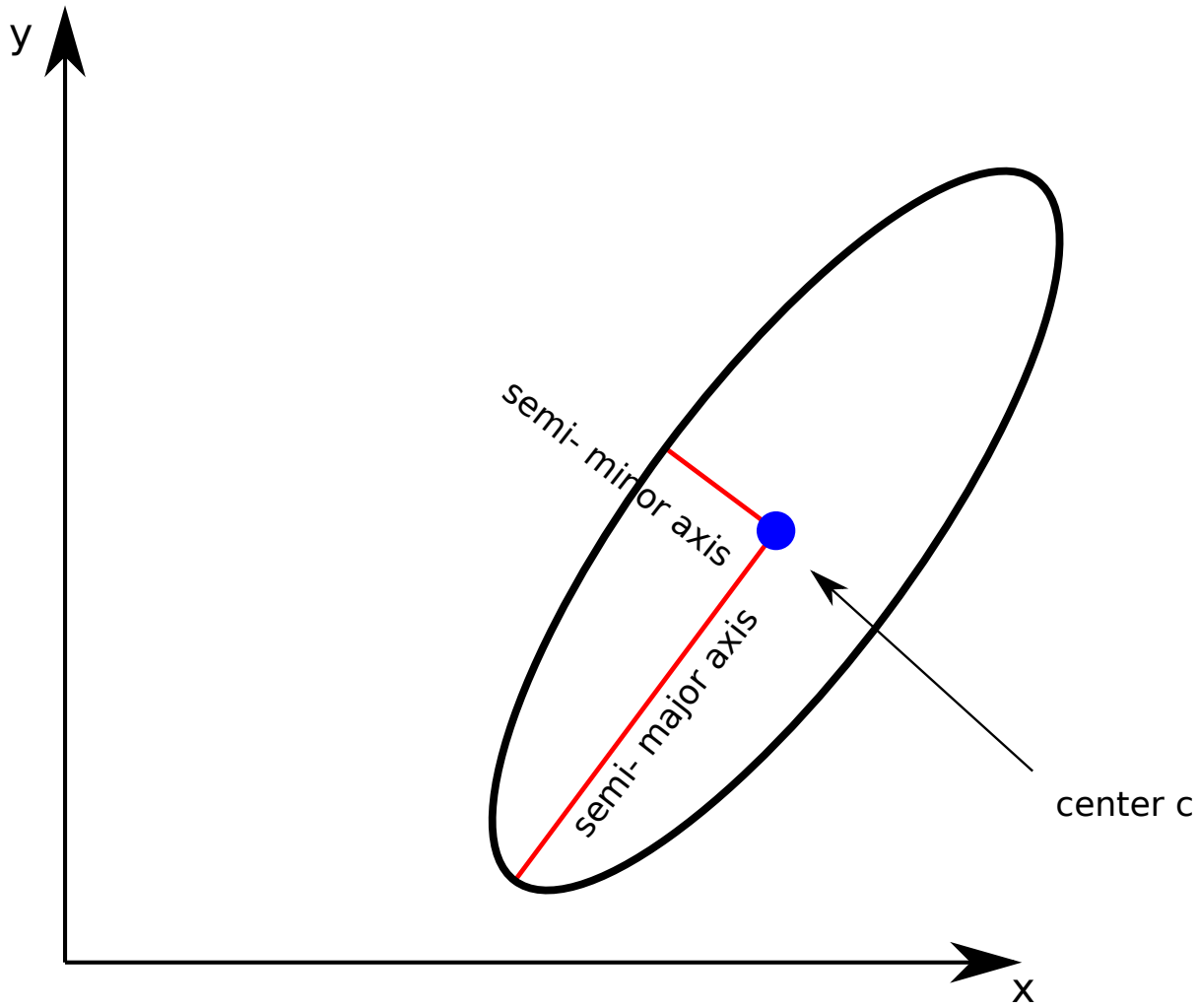
Introduction

What is an ellipse?



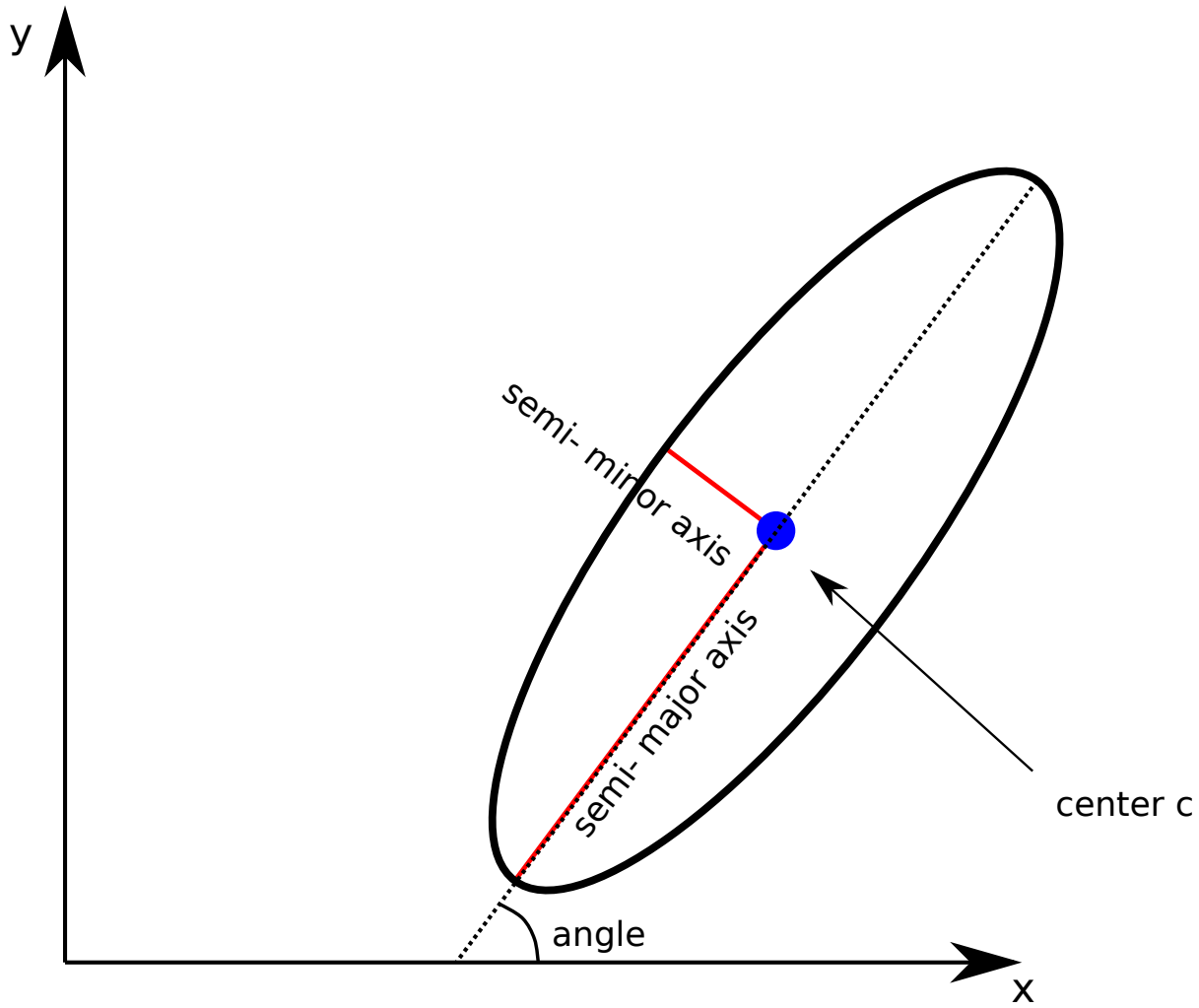
Introduction

What is an ellipse?

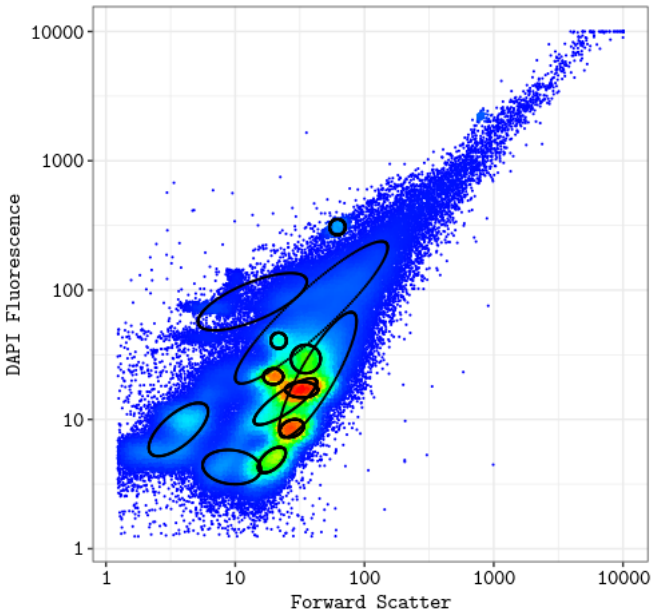


Introduction

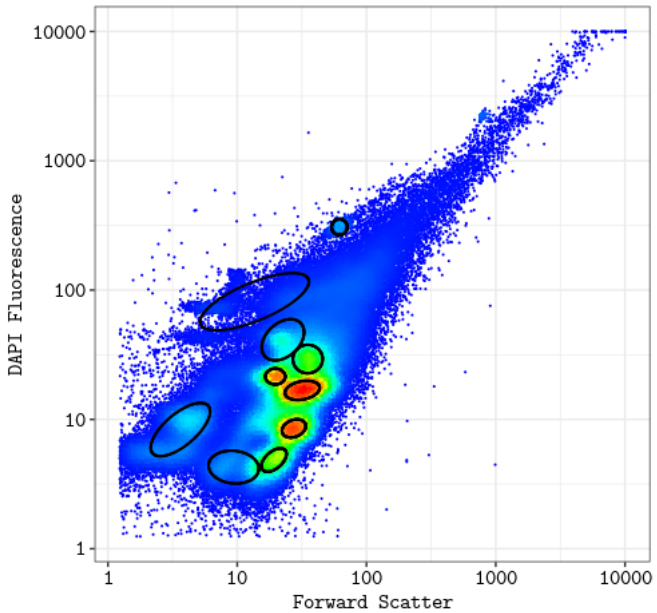
What is an ellipse?



Introduction



flowEMMi v1



flowEMMi v2

But how to remove the overlaps?

Merging

- Let E_1 and E_2 be two ellipses that overlap more than a given threshold,
  and let w_1 and w_2 be the weights of E_1 and E_2, respectively.

But how to remove the overlaps?

Merging

- Let E_1 and E_2 be two ellipses that overlap more than a given threshold,
  and let w_1 and w_2 be the weights of E_1 and E_2, respectively.
- Then, they will be "merged" into a new ellipse E_c. E_c will be more similar to the ellipse with the highest weight.

But how to remove the overlaps?

Merging

- Let E_1 and E_2 be two ellipses that overlap more than a given threshold,
  and let w_1 and w_2 be the weights of E_1 and E_2, respectively.
- Then, they will be "merged" into a new ellipse E_c. E_c will be more similar to the ellipse with the highest weight.
- Let Σ_1 and Σ_2 be the covariance matrices and let μ_1 and μ_2 be the centers of E_1 and E_2, respectively.

But how to remove the overlaps?

Merging

- Let E_1 and E_2 be two ellipses that overlap more than a given threshold,
  and let w_1 and w_2 be the weights of E_1 and E_2, respectively.
- Then, they will be "merged" into a new ellipse E_c. E_c will be more similar to the ellipse with the highest weight.
- Let Σ_1 and Σ_2 be the covariance matrices and let μ_1 and μ_2 be the centers of E_1 and E_2, respectively.
- E_c is the product of the two Gaussian densities, since we want the probability distribution of points lying in *both* ellipses.

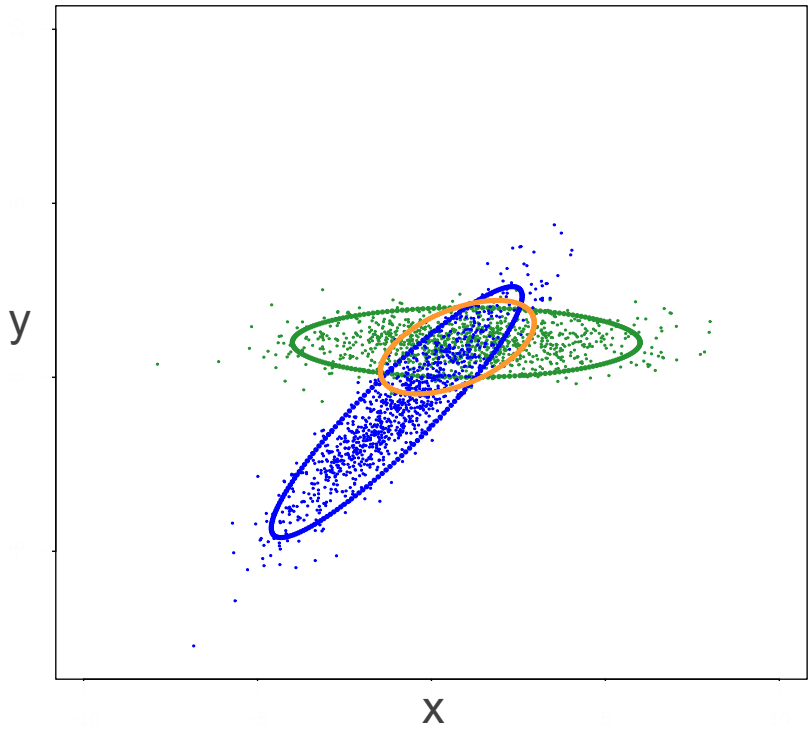
But how to remove the overlaps?

Merging

- Let E_1 and E_2 be two ellipses that overlap more than a given threshold,
  and let w_1 and w_2 be the weights of E_1 and E_2, respectively.
- Then, they will be "merged" into a new ellipse E_c. E_c will be more similar to the ellipse with the highest weight.
- Let Σ_1 and Σ_2 be the covariance matrices and let μ_1 and μ_2 be the centers of E_1 and E_2, respectively.
- E_c is the product of the two Gaussian densities, since we want the probability distribution of points lying in *both* ellipses.
- In the end, the original ellipses E_1 and E_2 will be deleted and replaced by E_c, i.e. by μ_c and Σ_c.

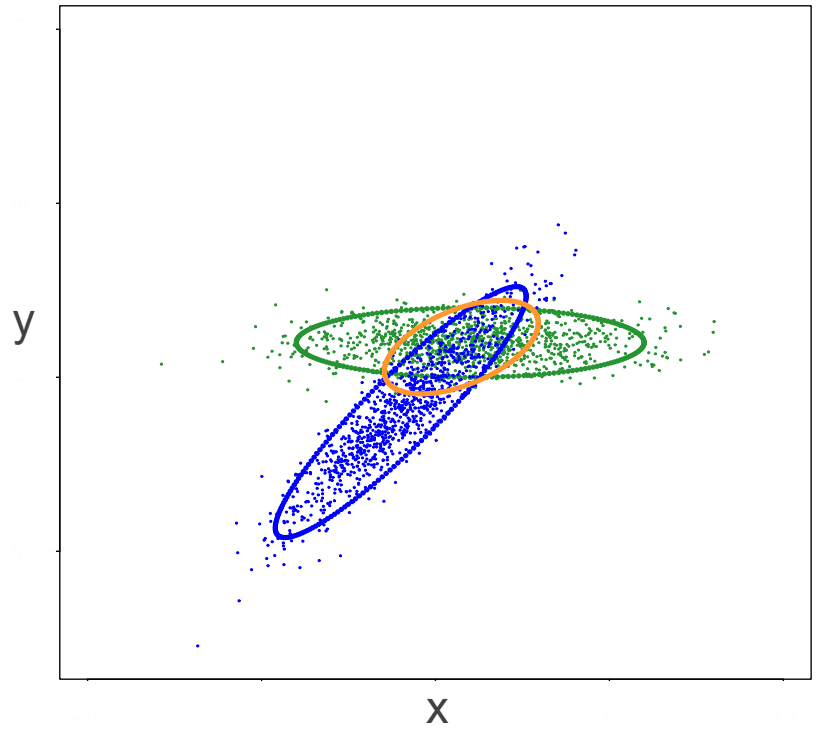
But how to remove the overlaps?

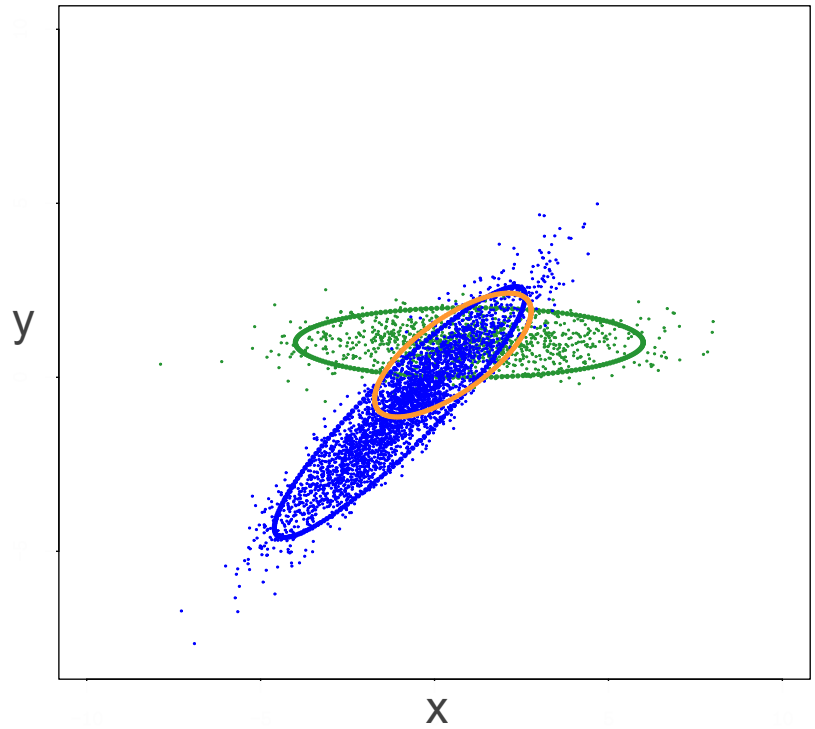
Merging



But how to remove the overlaps?

Merging

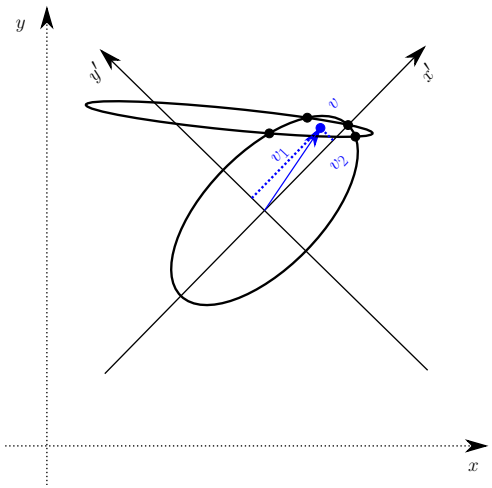




But how to remove the overlaps?

Shrinking

Let E_1 and E_2 be two ellipses that overlap with $o =$ size of intersection area.
Let w_1 and w_2 be the weights of E_1 and E_2 in the mixture model, respectively. Then:

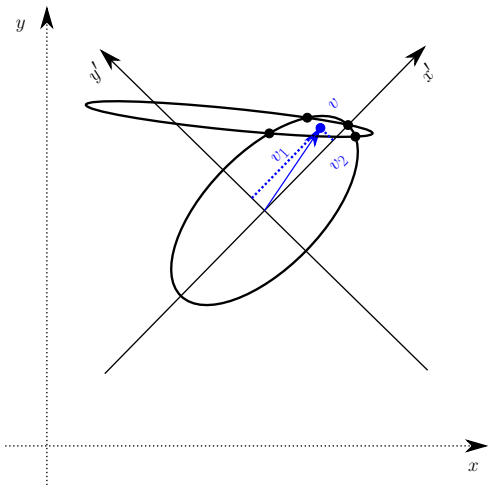


But how to remove the overlaps?

Shrinking

Let E_1 and E_2 be two ellipses that overlap with $o =$ size of intersection area.
Let w_1 and w_2 be the weights of E_1 and E_2 in the mixture model, respectively. Then:

1. Reduce the angle between the given ellipses by a given threshold.

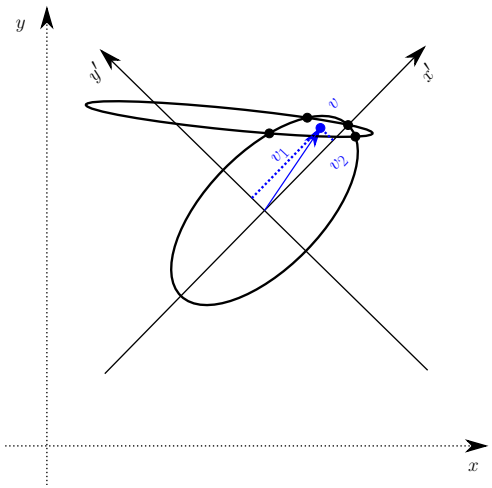


But how to remove the overlaps?

Shrinking

Let E_1 and E_2 be two ellipses that overlap with $o =$ size of intersection area.
Let w_1 and w_2 be the weights of E_1 and E_2 in the mixture model, respectively. Then:

1. Reduce the angle between the given ellipses by a given threshold.
2. Transform the coordinate system to the center of E_1 such that it is aligned with the semi-axes of E_1.

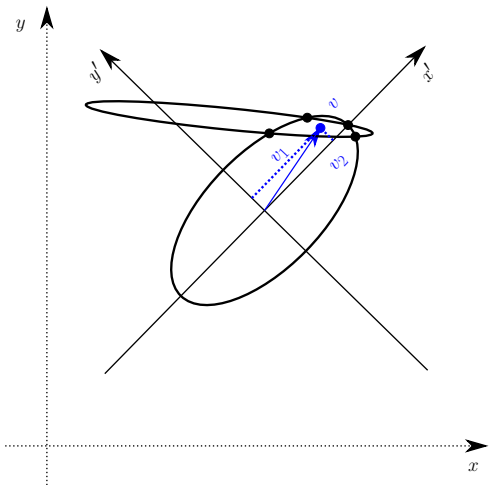


But how to remove the overlaps?

Shrinking

Let E_1 and E_2 be two ellipses that overlap with $o =$ size of intersection area.
Let w_1 and w_2 be the weights of E_1 and E_2 in the mixture model, respectively. Then:

1. Reduce the angle between the given ellipses by a given threshold.
2. Transform the coordinate system to the center of E_1 such that it is aligned with the semi-axes of E_1.
3. Determine the intersection points of E_1 and E_2.

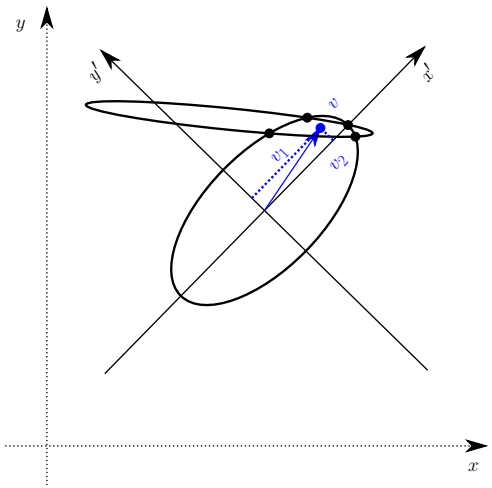


But how to remove the overlaps?

Shrinking

Let E_1 and E_2 be two ellipses that overlap with $o =$ size of intersection area.
Let w_1 and w_2 be the weights of E_1 and E_2 in the mixture model, respectively. Then:

1. Reduce the angle between the given ellipses by a given threshold.
2. Transform the coordinate system to the center of E_1 such that it is aligned with the semi-axes of E_1.
3. Determine the intersection points of E_1 and E_2.
4. Calculate the average of the intersection points; call this vector $v = (v_1, v_2)$.



But how to remove the overlaps?

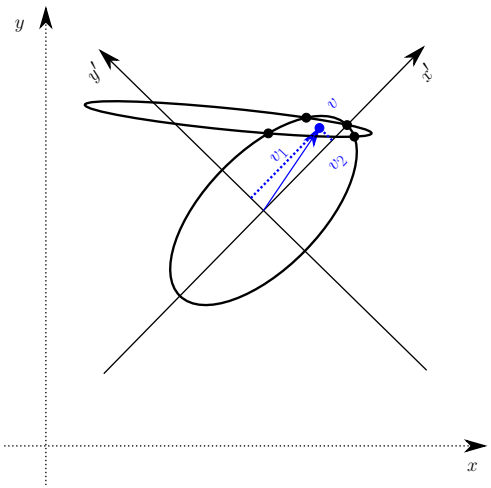
Shrinking

Let E_1 and E_2 be two ellipses that overlap with $o =$ size of intersection area.
Let w_1 and w_2 be the weights of E_1 and E_2 in the mixture model, respectively. Then:

1. Reduce the angle between the given ellipses by a given threshold.
2. Transform the coordinate system to the center of E_1 such that it is aligned with the semi-axes of E_1.
3. Determine the intersection points of E_1 and E_2.
4. Calculate the average of the intersection points; call this vector $v = (v_1, v_2)$.
5. Recalculate E_1:

$$\text{major}_{\text{new}} = \text{major} \cdot \left(1 - \sqrt{\frac{o}{|E_1|} \cdot \frac{|v_1|}{|v_1| + |v_2|} \cdot \frac{w_2}{w_1 + w_2}}\right)$$

$$\text{minor}_{\text{new}} = \text{minor} \cdot \left(1 - \sqrt{\frac{o}{|E_1|} \cdot \frac{|v_2|}{|v_1| + |v_2|} \cdot \frac{w_2}{w_1 + w_2}}\right)$$



But how to remove the overlaps?

Shrinking

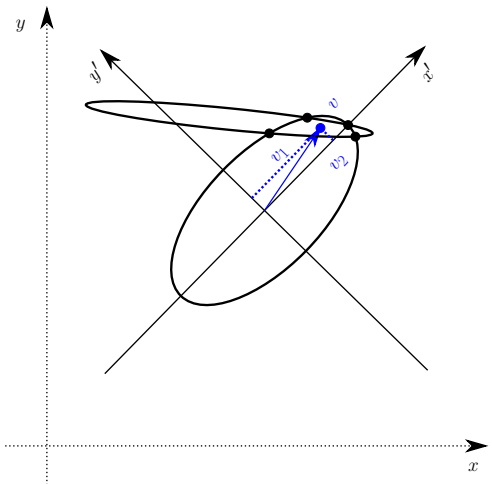
Let E_1 and E_2 be two ellipses that overlap with $o =$ size of intersection area.
Let w_1 and w_2 be the weights of E_1 and E_2 in the mixture model, respectively. Then:

1. Reduce the angle between the given ellipses by a given threshold.
2. Transform the coordinate system to the center of E_1 such that it is aligned with the semi-axes of E_1.
3. Determine the intersection points of E_1 and E_2.
4. Calculate the average of the intersection points; call this vector $v = (v_1, v_2)$.
5. Recalculate E_1:

$$\text{major}_{\text{new}} = \text{major} \cdot \left(1 - \sqrt{\frac{o}{|E_1|} \cdot \frac{|v_1|}{|v_1| + |v_2|} \cdot \frac{w_2}{w_1 + w_2}}\right)$$

$$\text{minor}_{\text{new}} = \text{minor} \cdot \left(1 - \sqrt{\frac{o}{|E_1|} \cdot \frac{|v_2|}{|v_1| + |v_2|} \cdot \frac{w_2}{w_1 + w_2}}\right)$$

6. Recalculate E_2 in the same way.



But how to remove the overlaps?

Shrinking

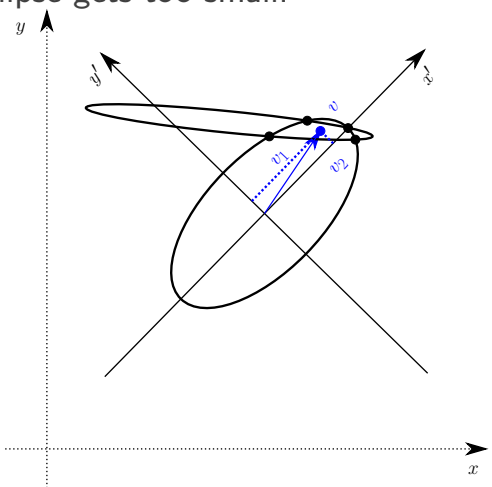
Let E_1 and E_2 be two ellipses that overlap with $o =$ size of intersection area.
Let w_1 and w_2 be the weights of E_1 and E_2 in the mixture model, respectively. Then:

1. Reduce the angle between the given ellipses by a given threshold.
2. Transform the coordinate system to the center of E_1 such that it is aligned with the semi-axes of E_1.
3. Determine the intersection points of E_1 and E_2.
4. Calculate the average of the intersection points; call this vector $v = (v_1, v_2)$.
5. Recalculate E_1:

$$\text{major}_{\text{new}} = \text{major} \cdot \left(1 - \sqrt{\frac{o}{|E_1|} \cdot \frac{|v_1|}{|v_1| + |v_2|} \cdot \frac{w_2}{w_1 + w_2}}\right)$$

$$\text{minor}_{\text{new}} = \text{minor} \cdot \left(1 - \sqrt{\frac{o}{|E_1|} \cdot \frac{|v_2|}{|v_1| + |v_2|} \cdot \frac{w_2}{w_1 + w_2}}\right)$$

6. Recalculate E_2 in the same way.
7. Repeat steps 1. to 6. until E_1 and E_2 do not overlap anymore or until one ellipse gets too small.



But how to remove the overlaps?

Shrinking

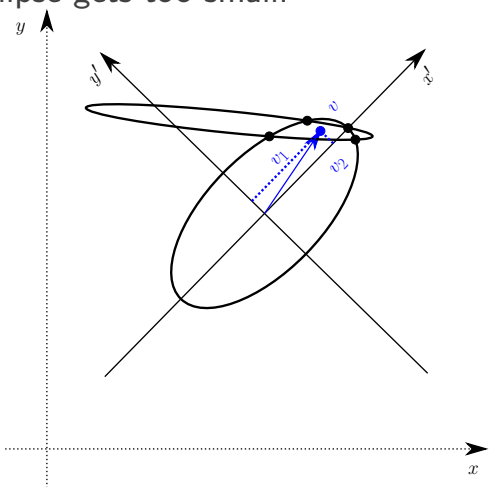
Let E_1 and E_2 be two ellipses that overlap with $o =$ size of intersection area.
Let w_1 and w_2 be the weights of E_1 and E_2 in the mixture model, respectively. Then:

1. Reduce the angle between the given ellipses by a given threshold.
2. Transform the coordinate system to the center of E_1 such that it is aligned with the semi-axes of E_1.
3. Determine the intersection points of E_1 and E_2.
4. Calculate the average of the intersection points; call this vector $v = (v_1, v_2)$.
5. Recalculate E_1:

$$\text{major}_{\text{new}} = \text{major} \cdot \left(1 - \sqrt{\frac{o}{|E_1|} \cdot \frac{|v_1|}{|v_1| + |v_2|} \cdot \frac{w_2}{w_1 + w_2}}\right)$$

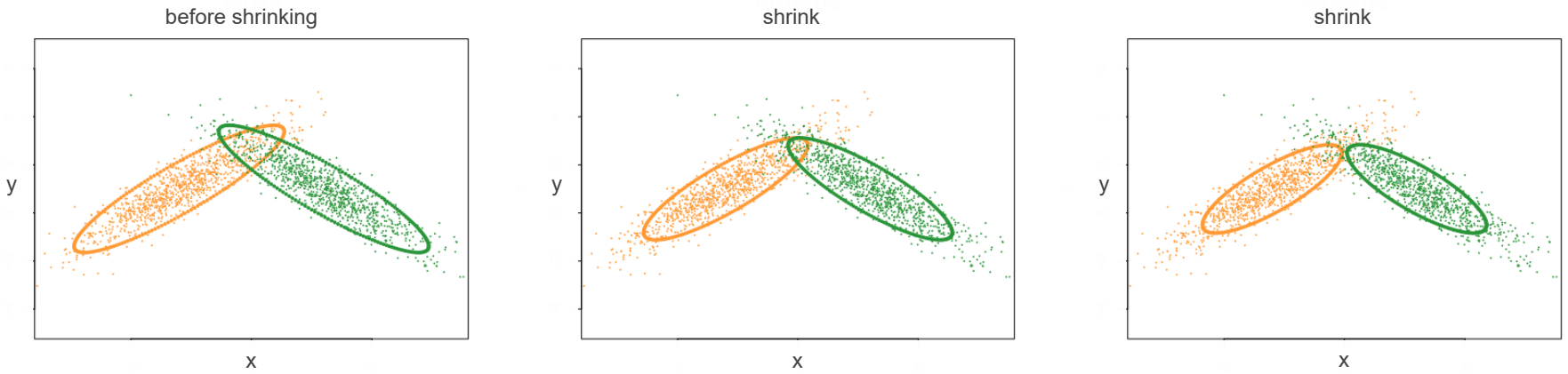
$$\text{minor}_{\text{new}} = \text{minor} \cdot \left(1 - \sqrt{\frac{o}{|E_1|} \cdot \frac{|v_2|}{|v_1| + |v_2|} \cdot \frac{w_2}{w_1 + w_2}}\right)$$

6. Recalculate E_2 in the same way.
7. Repeat steps 1. to 6. until E_1 and E_2 do not overlap anymore or until one ellipse gets too small.
8. If one ellipse gets too small, then merge the original E_1 and E_2 instead.



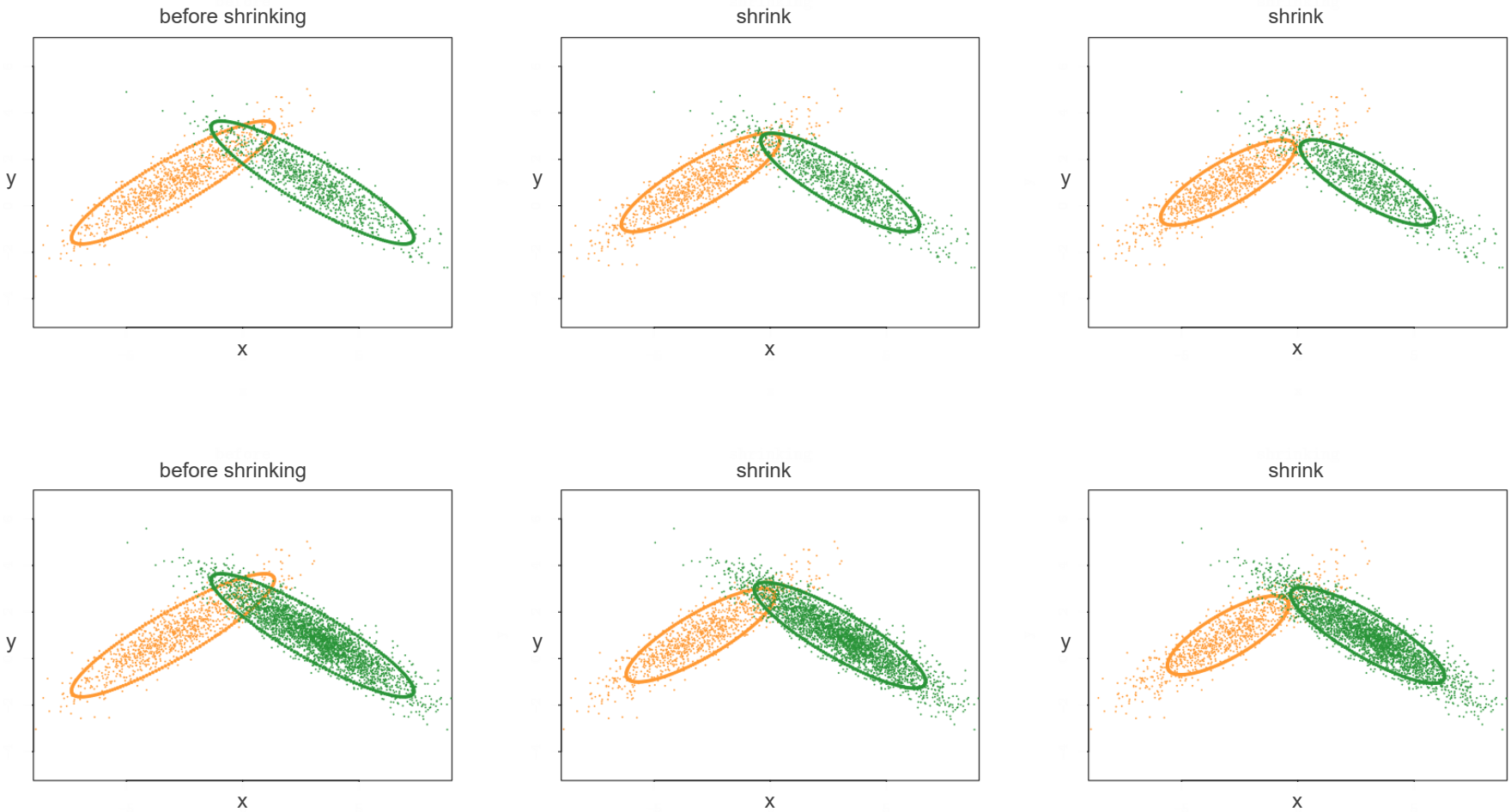
But how to remove the overlaps?

Shrinking



But how to remove the overlaps?

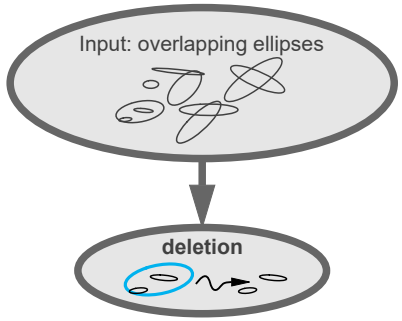
Shrinking



Workflow

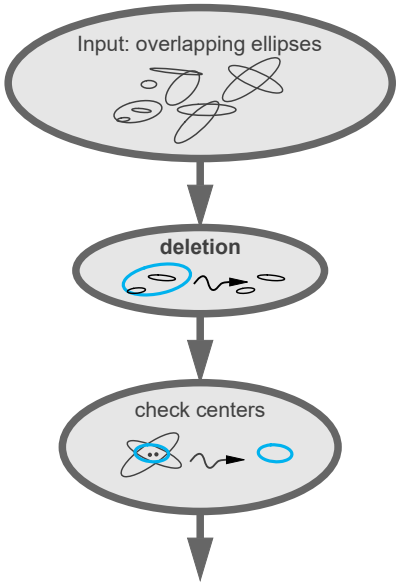
HELMHOLTZ Zentrum für Umweltforschung

Workflow



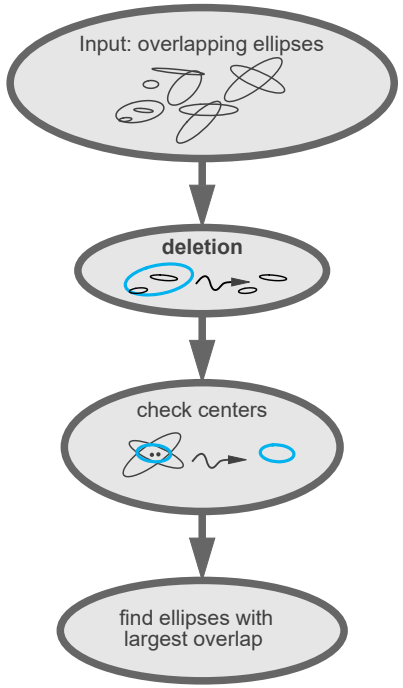
HELMHOLTZ Zentrum für Umweltforschung

Workflow

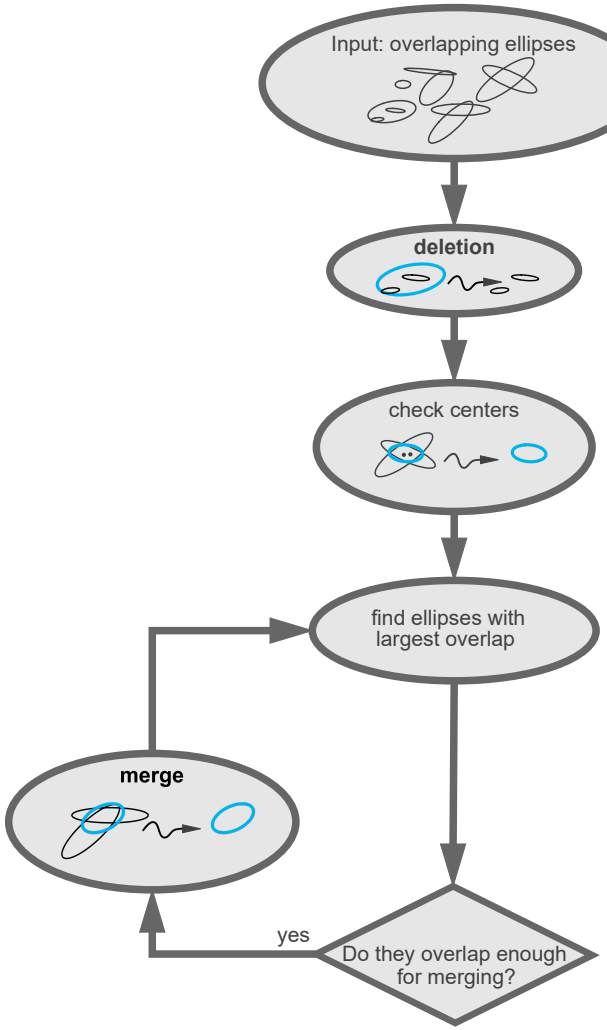


HELMHOLTZ Zentrum für Umweltforschung

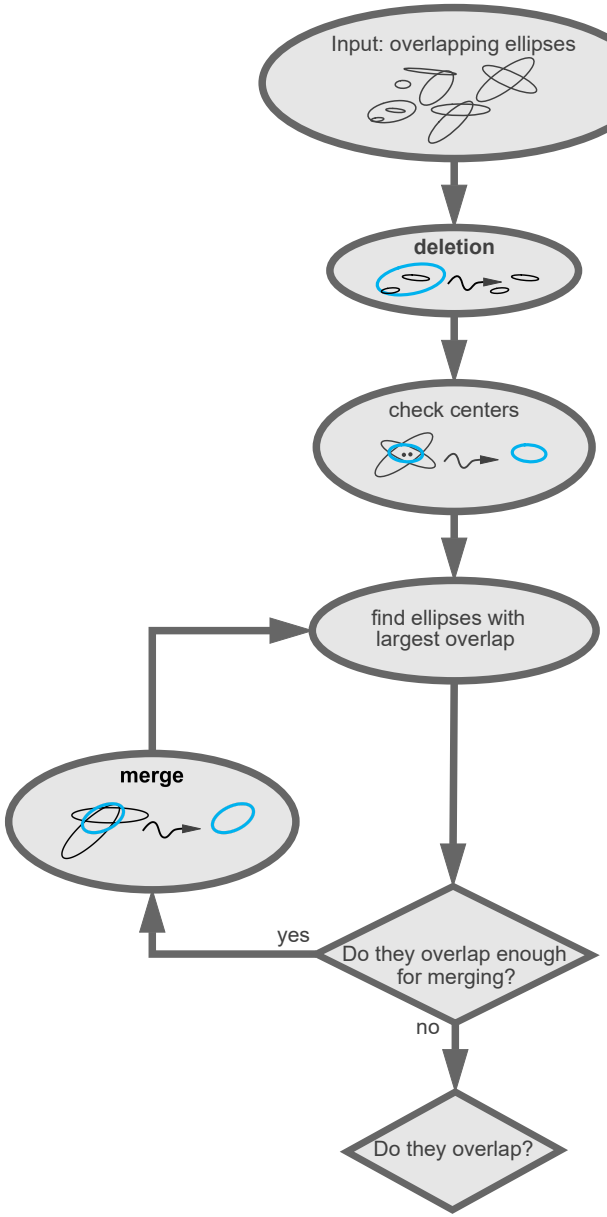
Workflow


Input: overlapping ellipses
deletion
check centers
find ellipses with
largest overlap

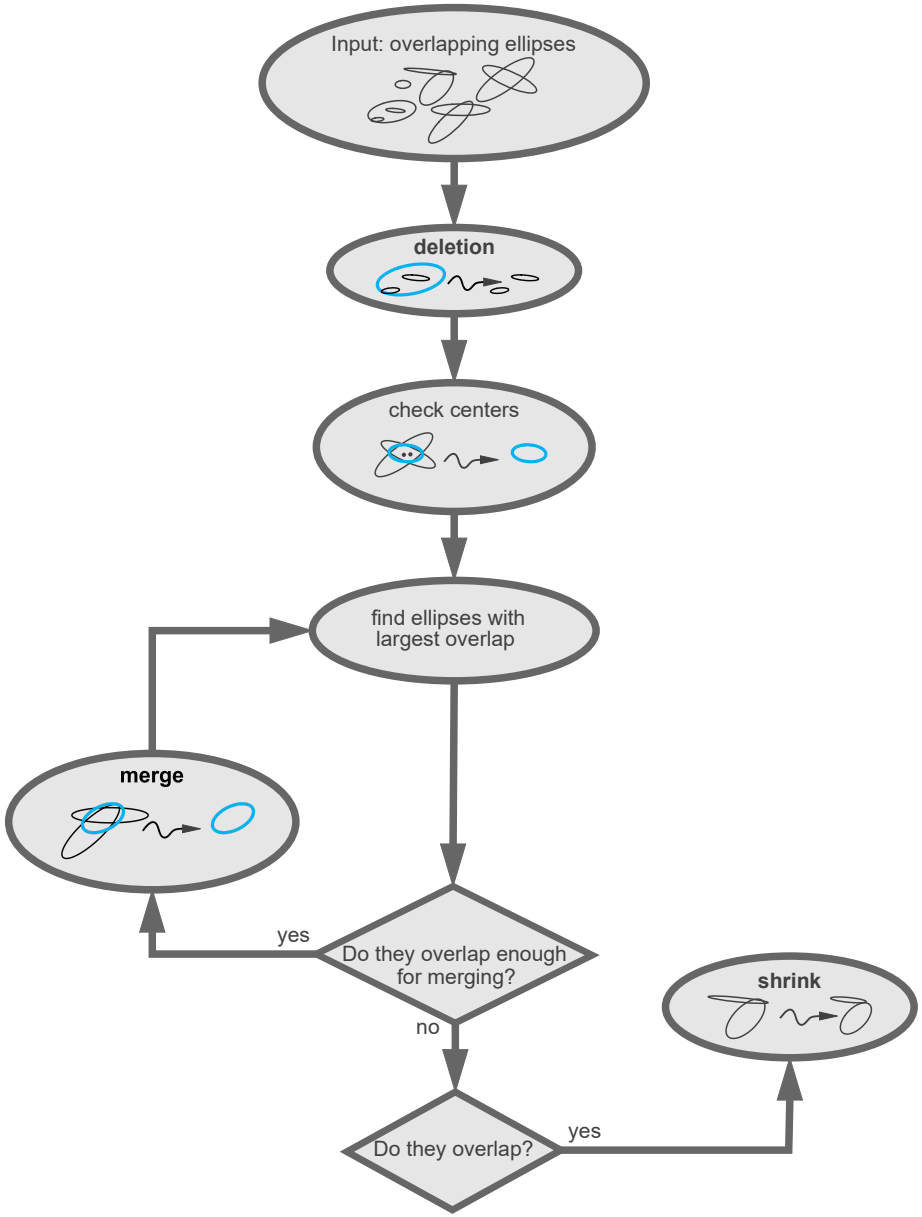
Workflow


Input: overlapping ellipses
deletion
check centers
find ellipses with largest overlap
merge
yes
Do they overlap enough for merging?

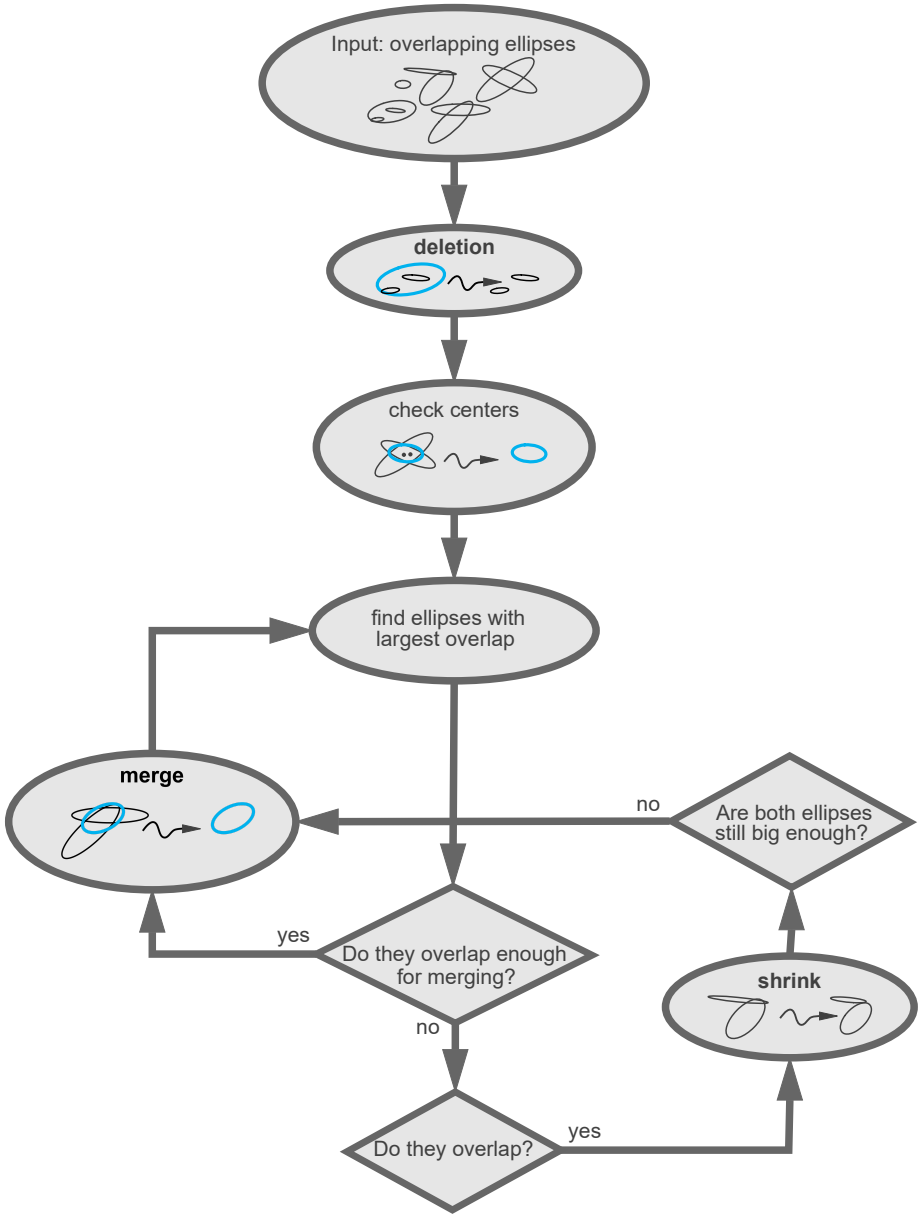
Workflow


Input: overlapping ellipses
deletion
check centers
find ellipses with largest overlap
merge
yes
Do they overlap enough for merging?
no
Do they overlap?

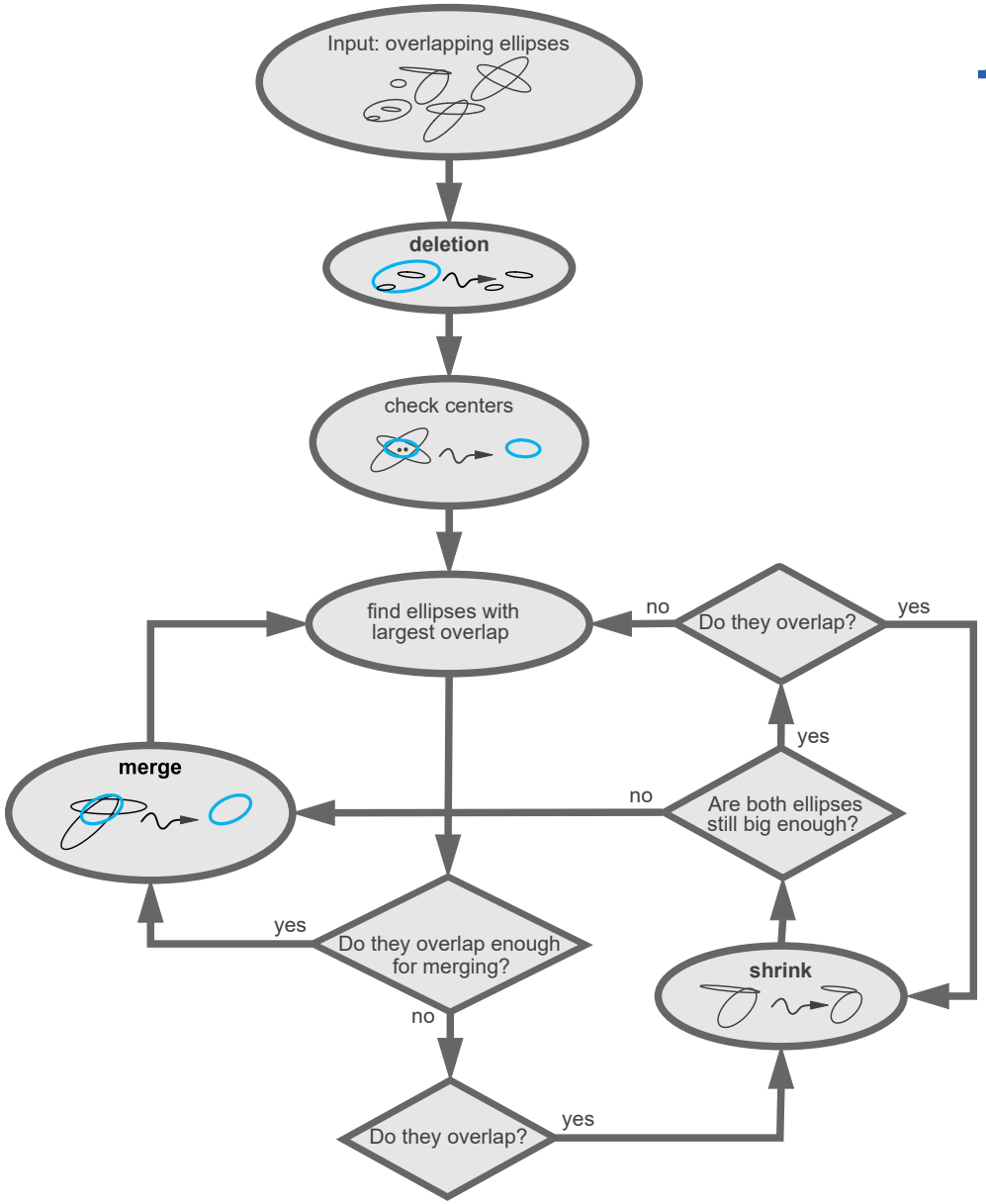
Workflow


Input: overlapping ellipses
deletion
check centers
find ellipses with largest overlap
merge
yes
Do they overlap enough for merging?
shrink
no
Do they overlap?
yes

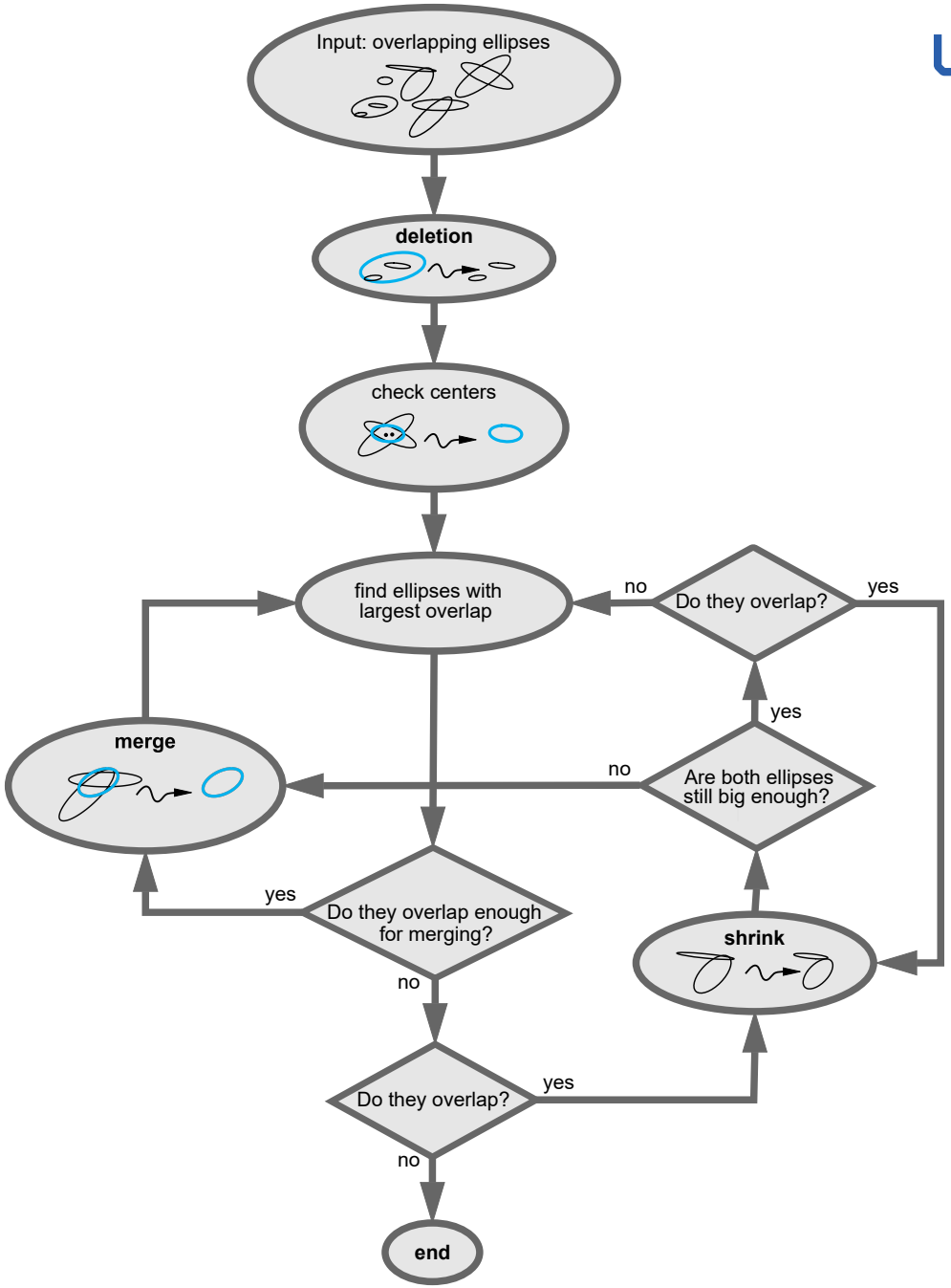
Workflow


Input: overlapping ellipses
deletion
check centers
find ellipses with largest overlap
merge
no
Are both ellipses still big enough?
yes
Do they overlap enough for merging?
shrink
no
Do they overlap?
yes

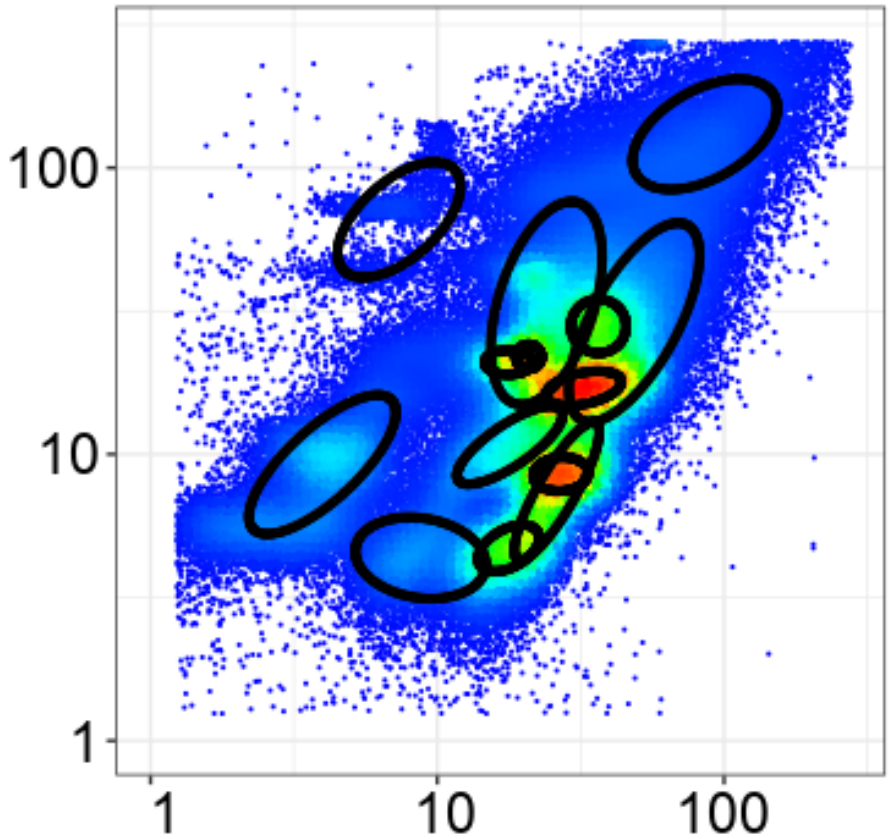
Workflow


Input: overlapping ellipses
deletion
check centers
find ellipses with largest overlap
Do they overlap?
no
yes
merge
Are both ellipses still big enough?
yes
no
Do they overlap enough for merging?
yes
shrink
no
Do they overlap?
yes

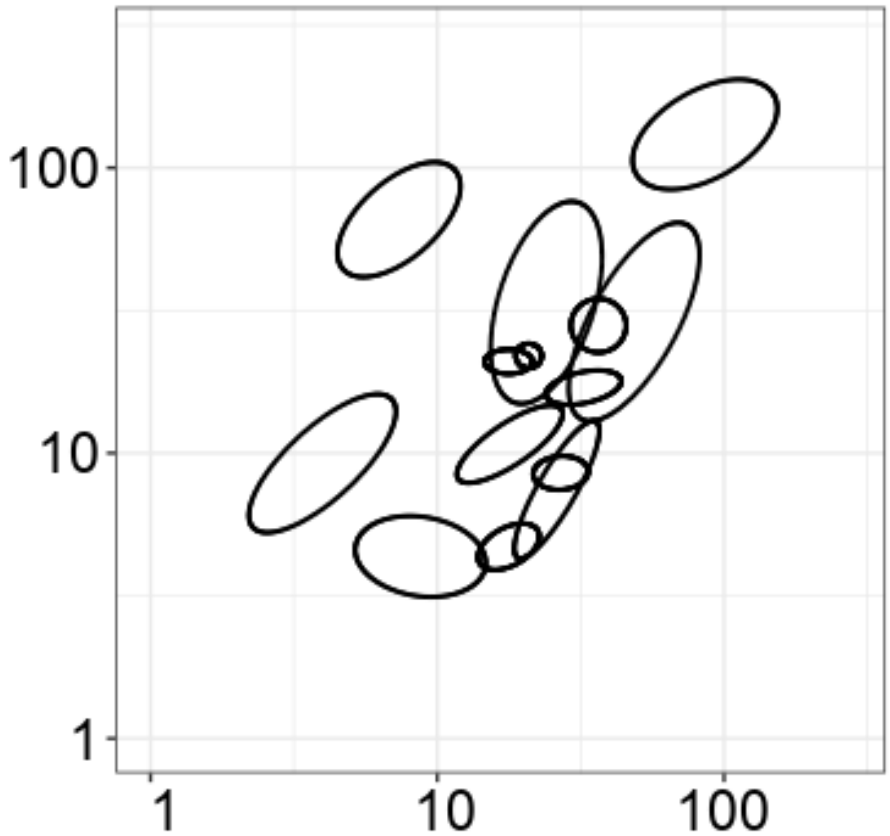
Workflow


Input: overlapping ellipses
deletion
check centers
find ellipses with largest overlap
Do they overlap?
no
yes
merge
Are both ellipses still big enough?
yes
no
Do they overlap enough for merging?
yes
shrink
no
Do they overlap?
yes
no
end

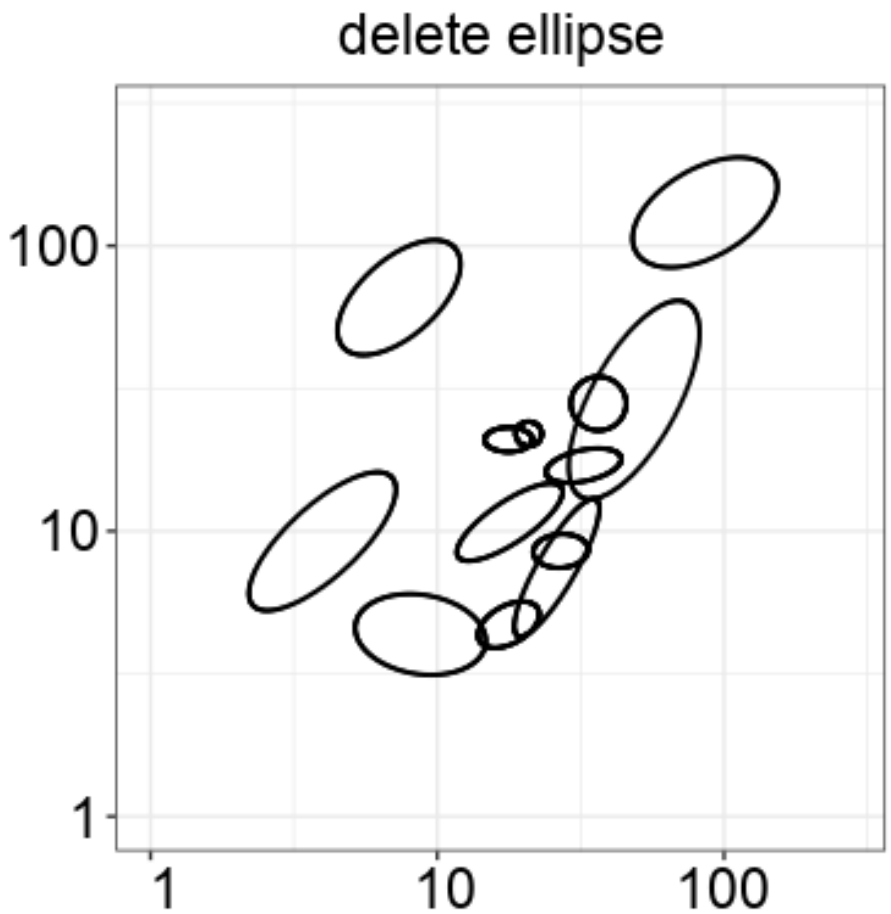
Example



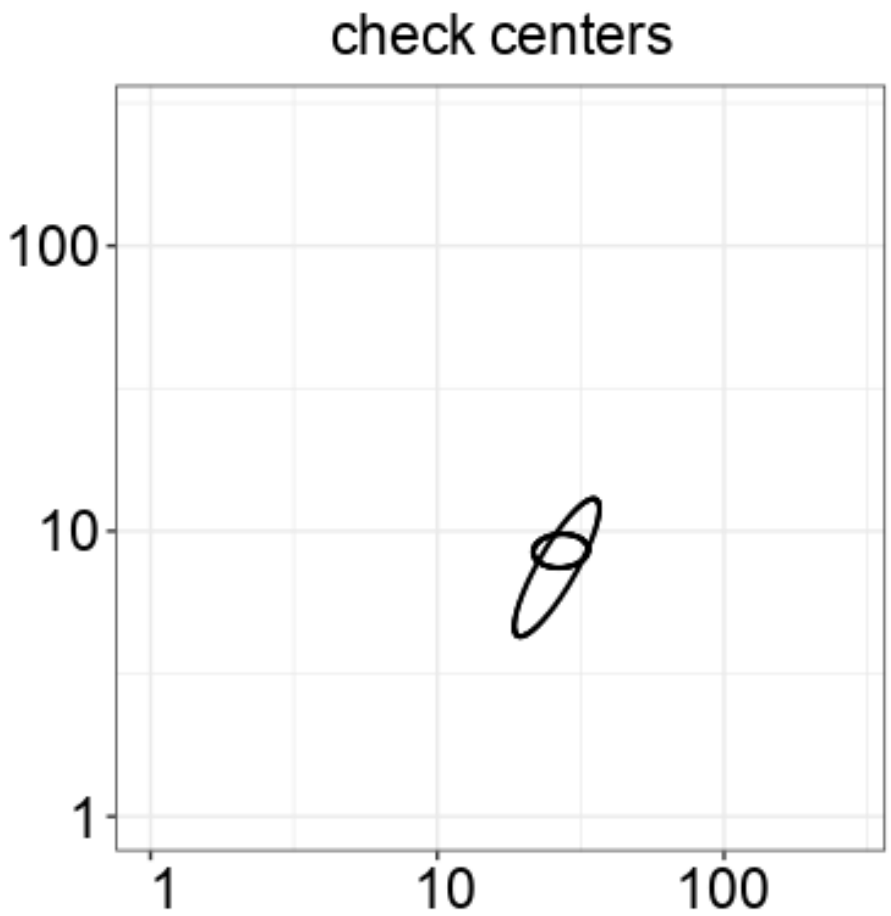
Example



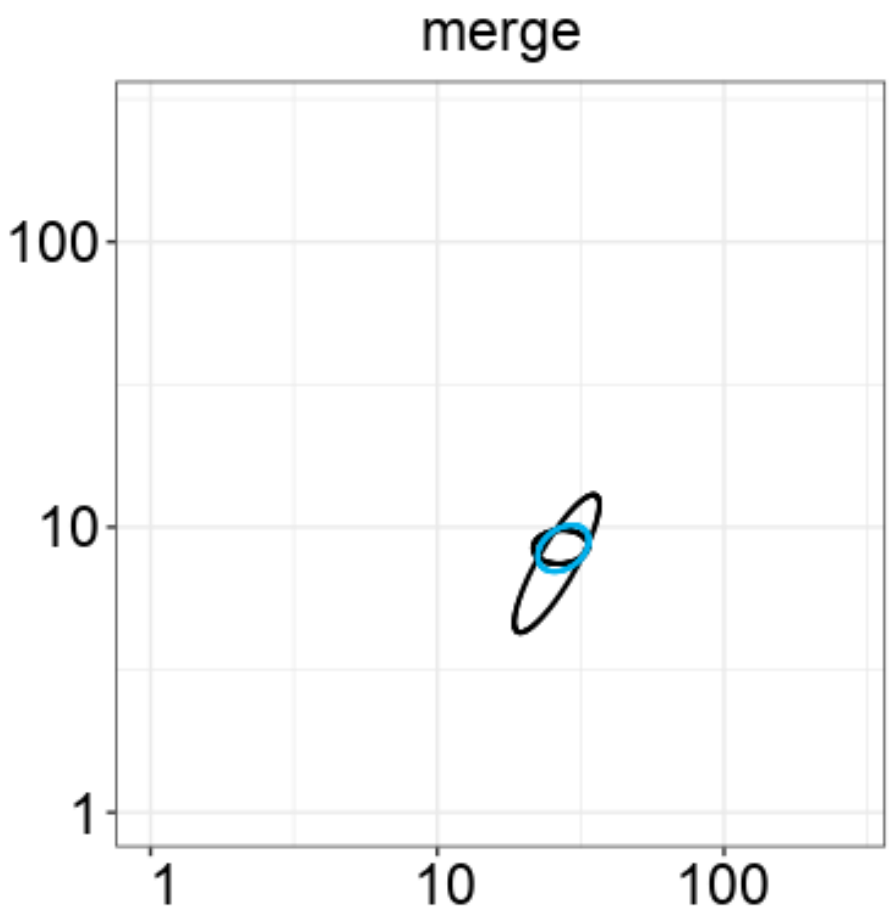
Example



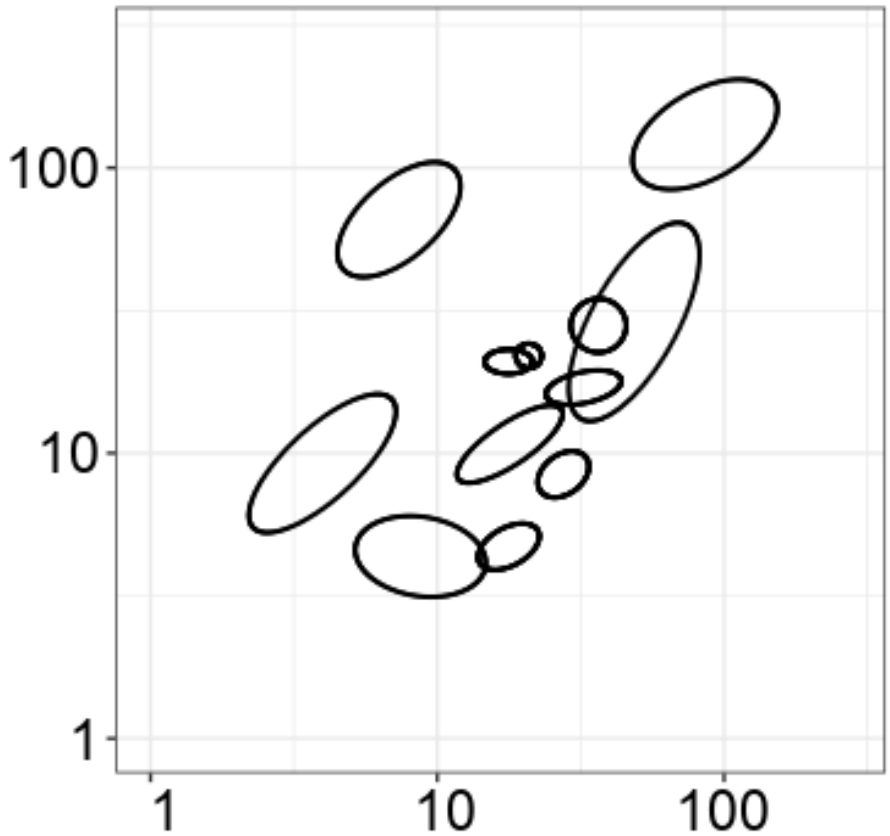
Example



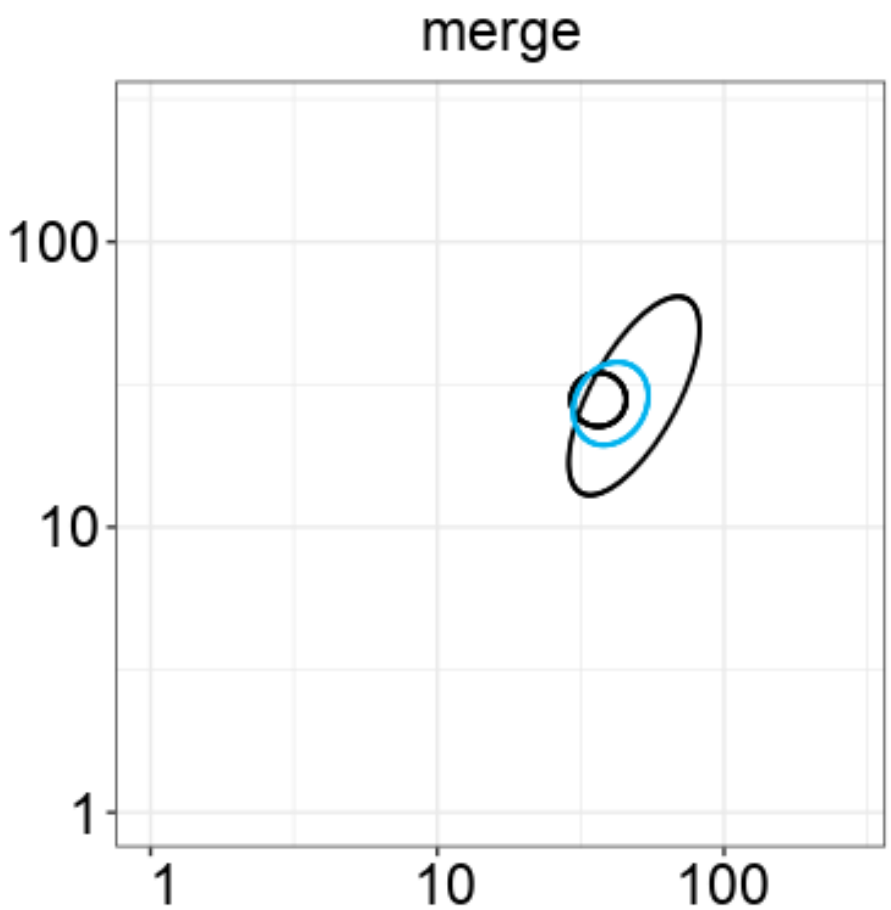
Example



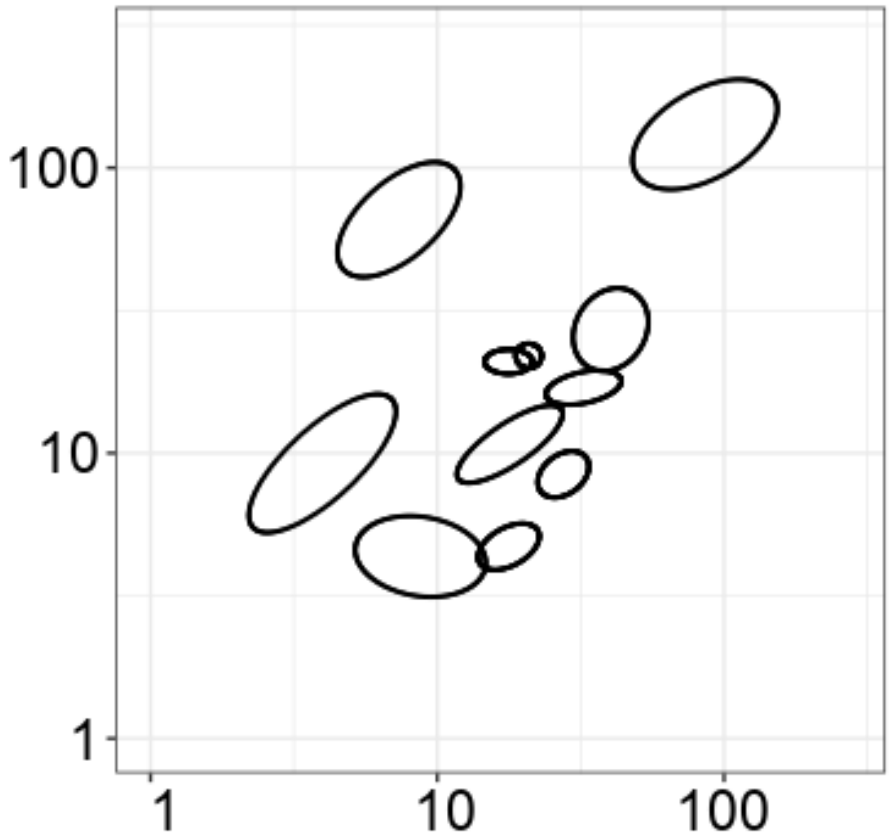
Example



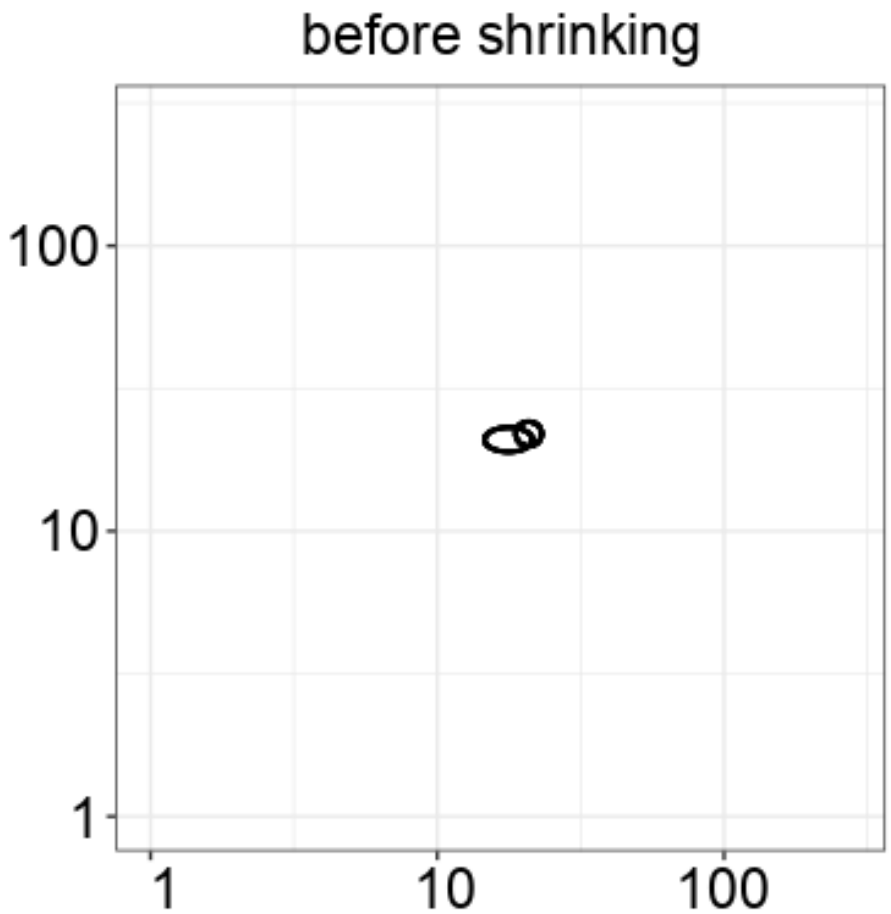
Example



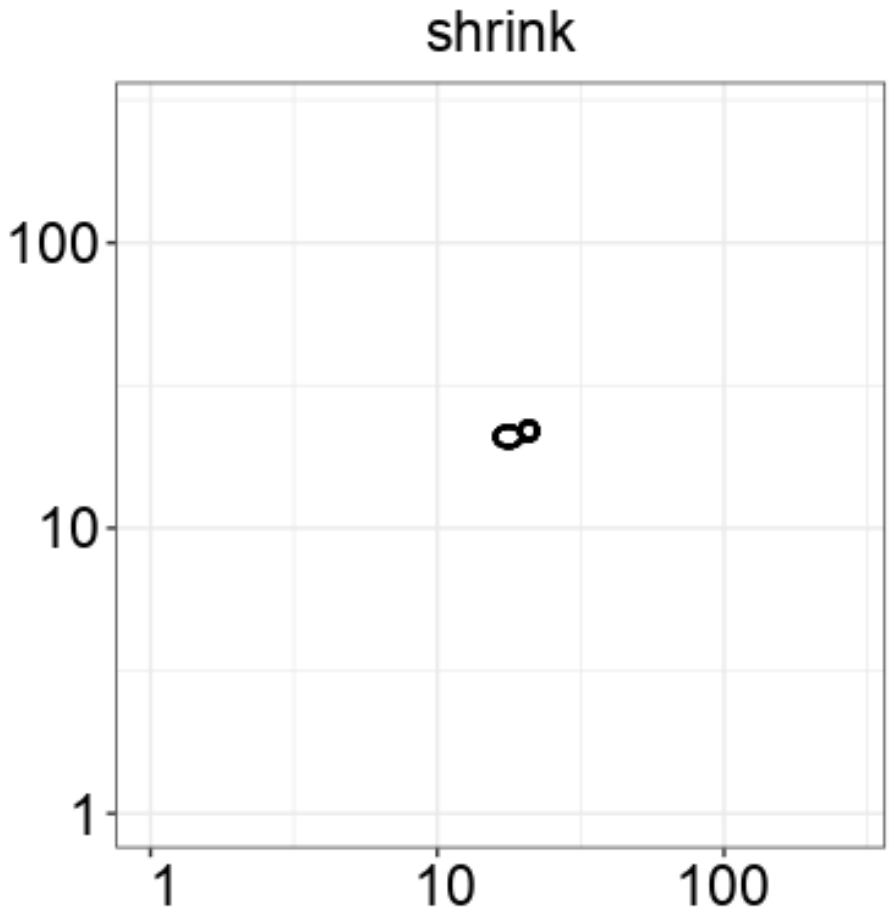
Example



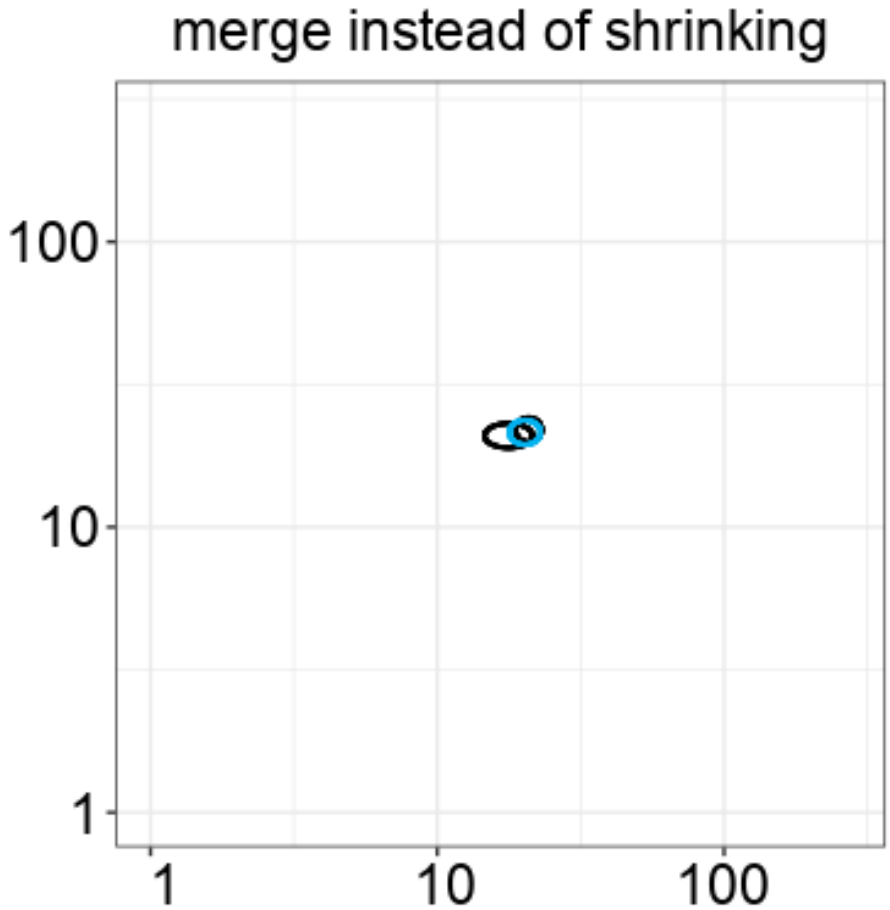
Example



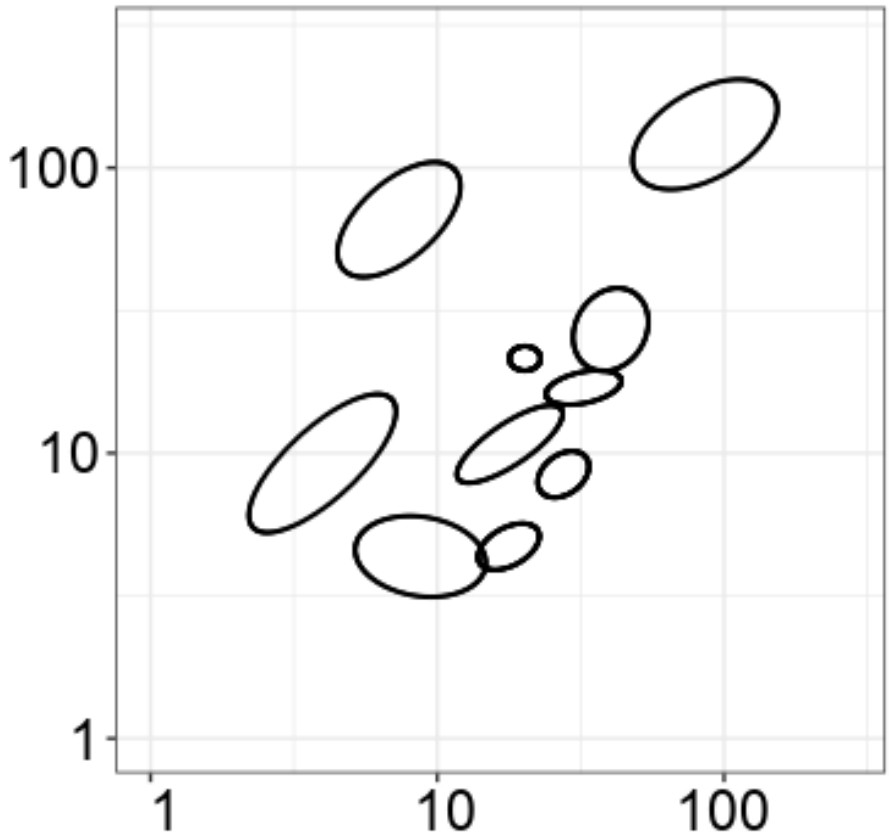
Example



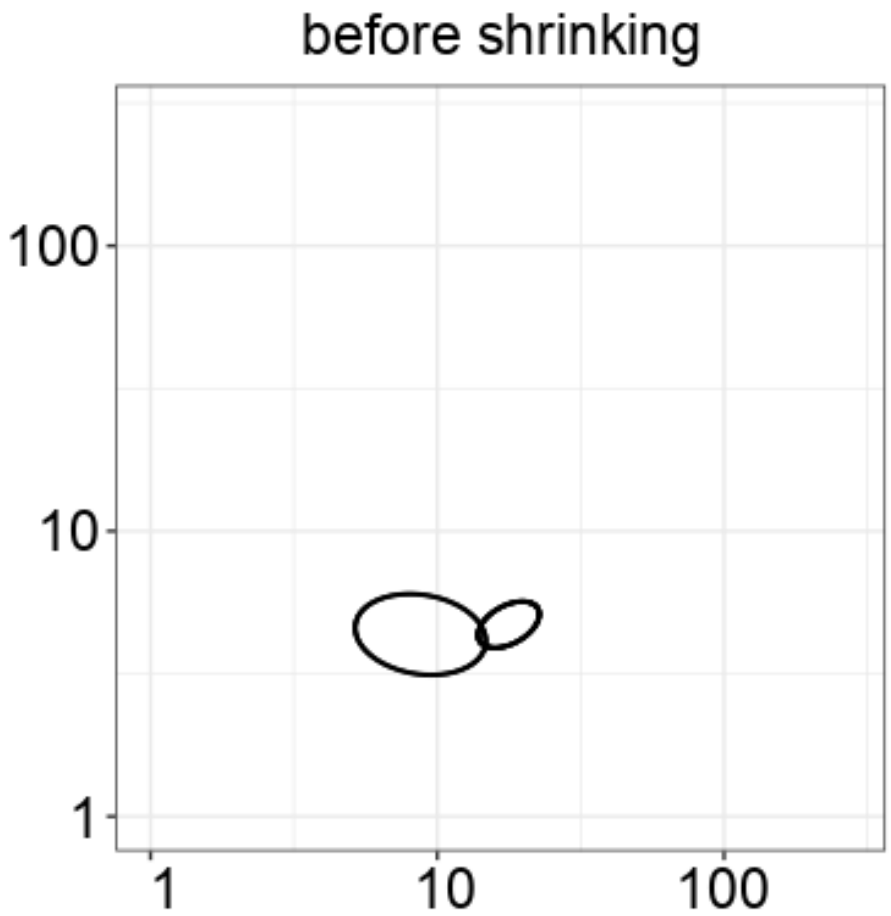
Example



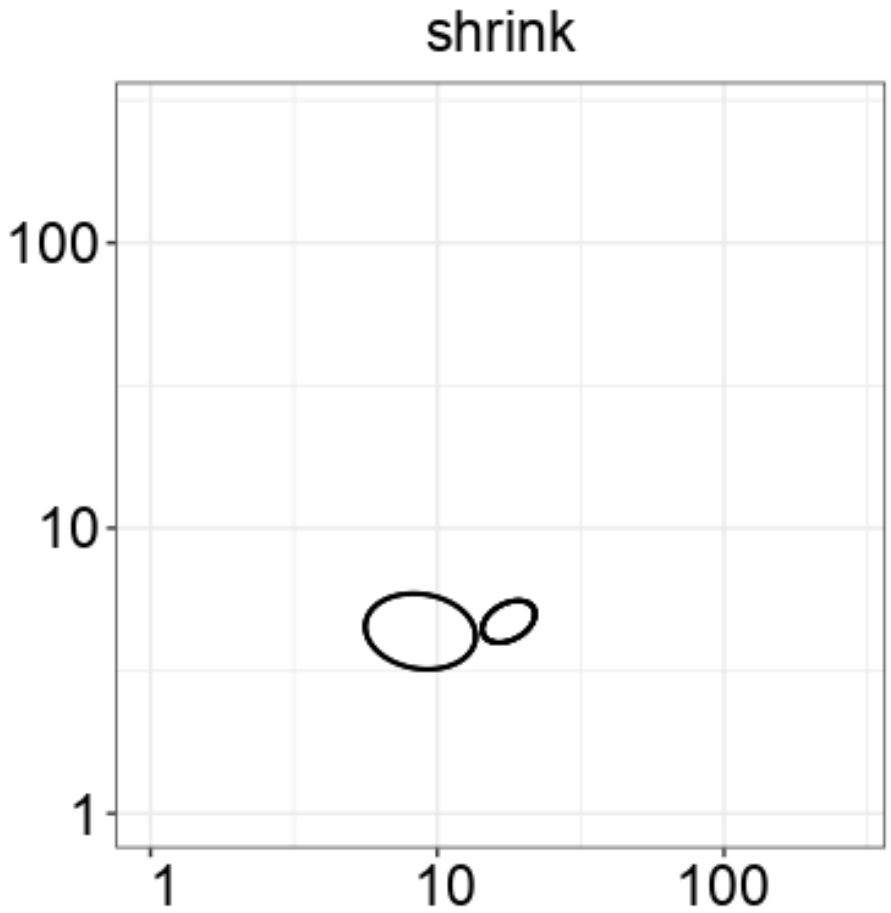
Example



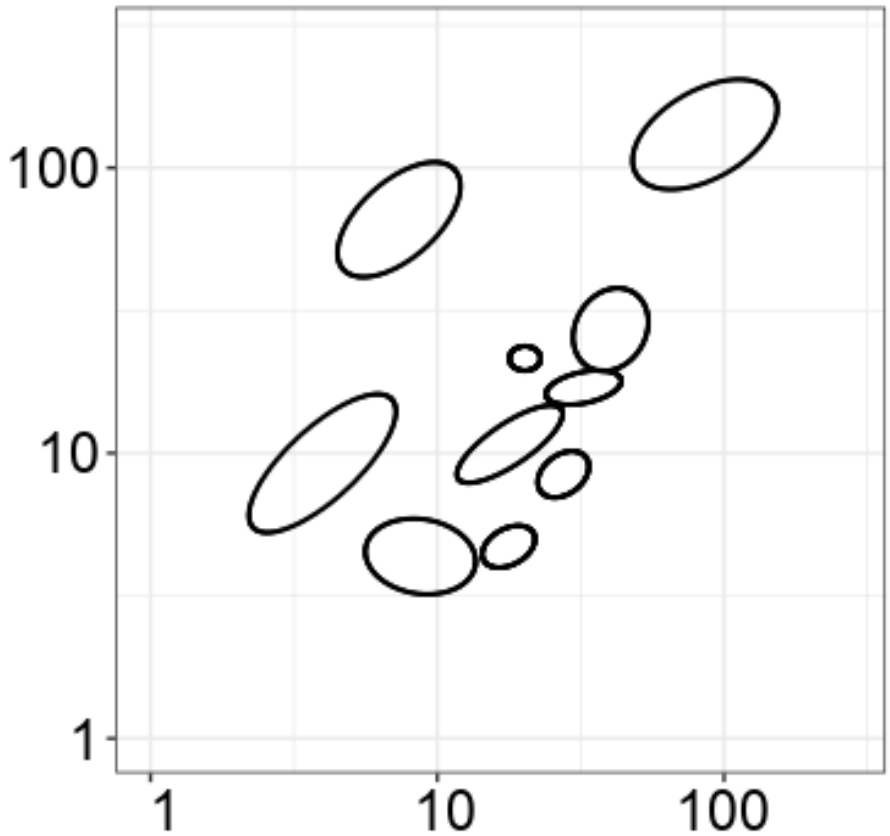
Example



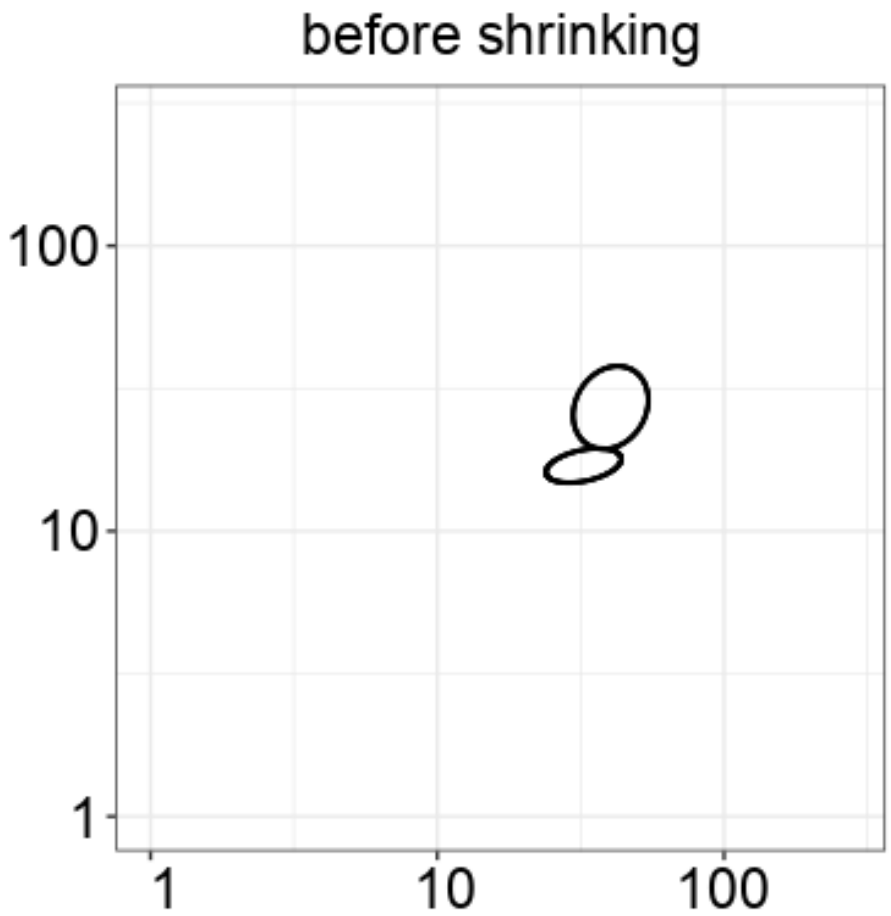
Example



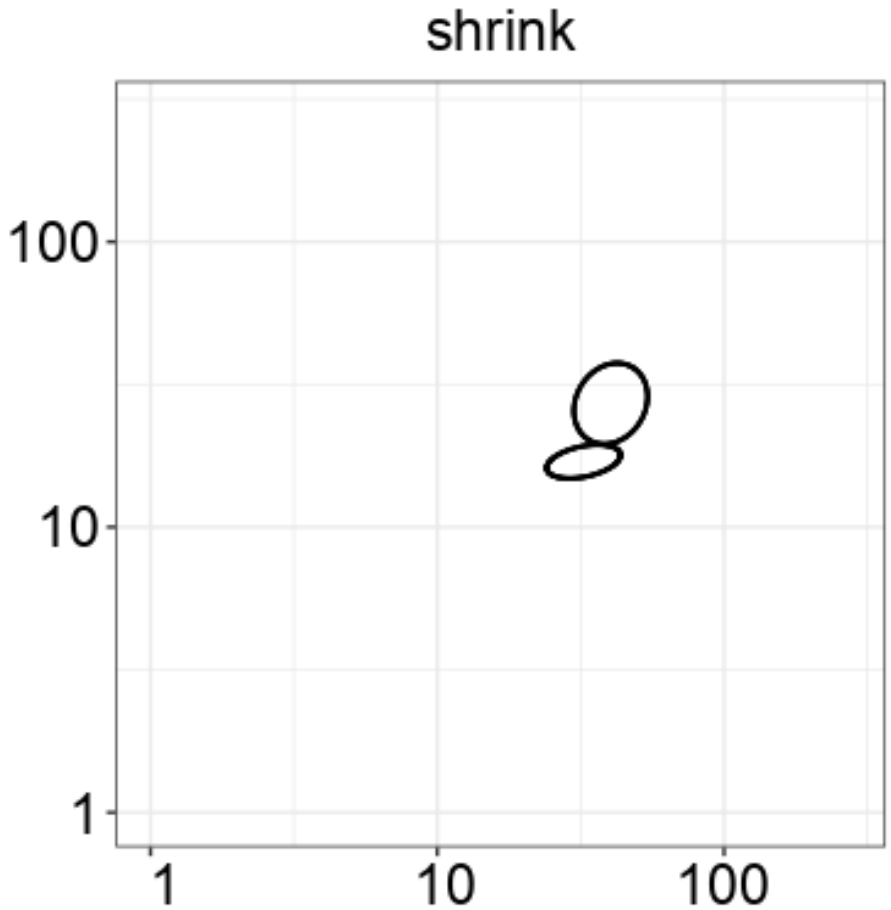
Example



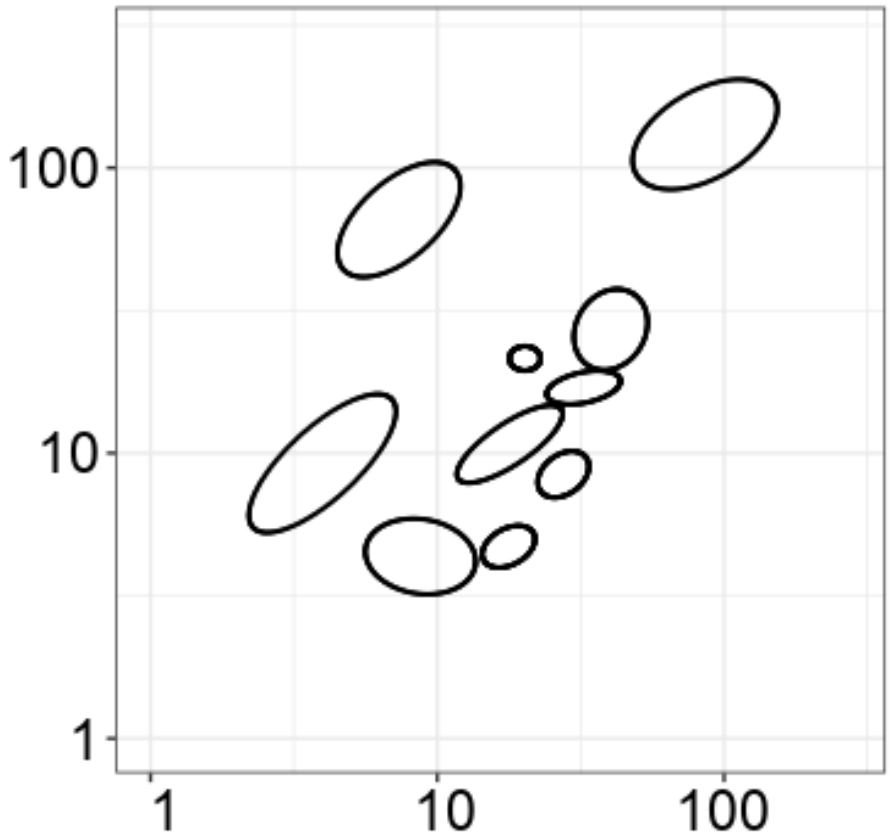
Example



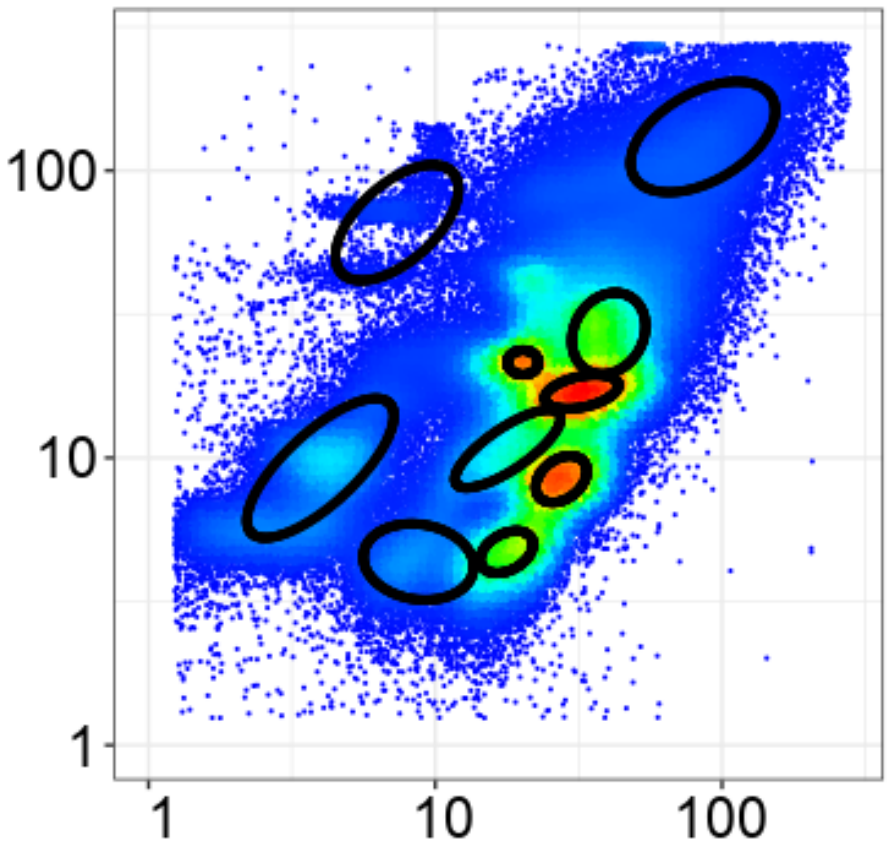
Example



Example

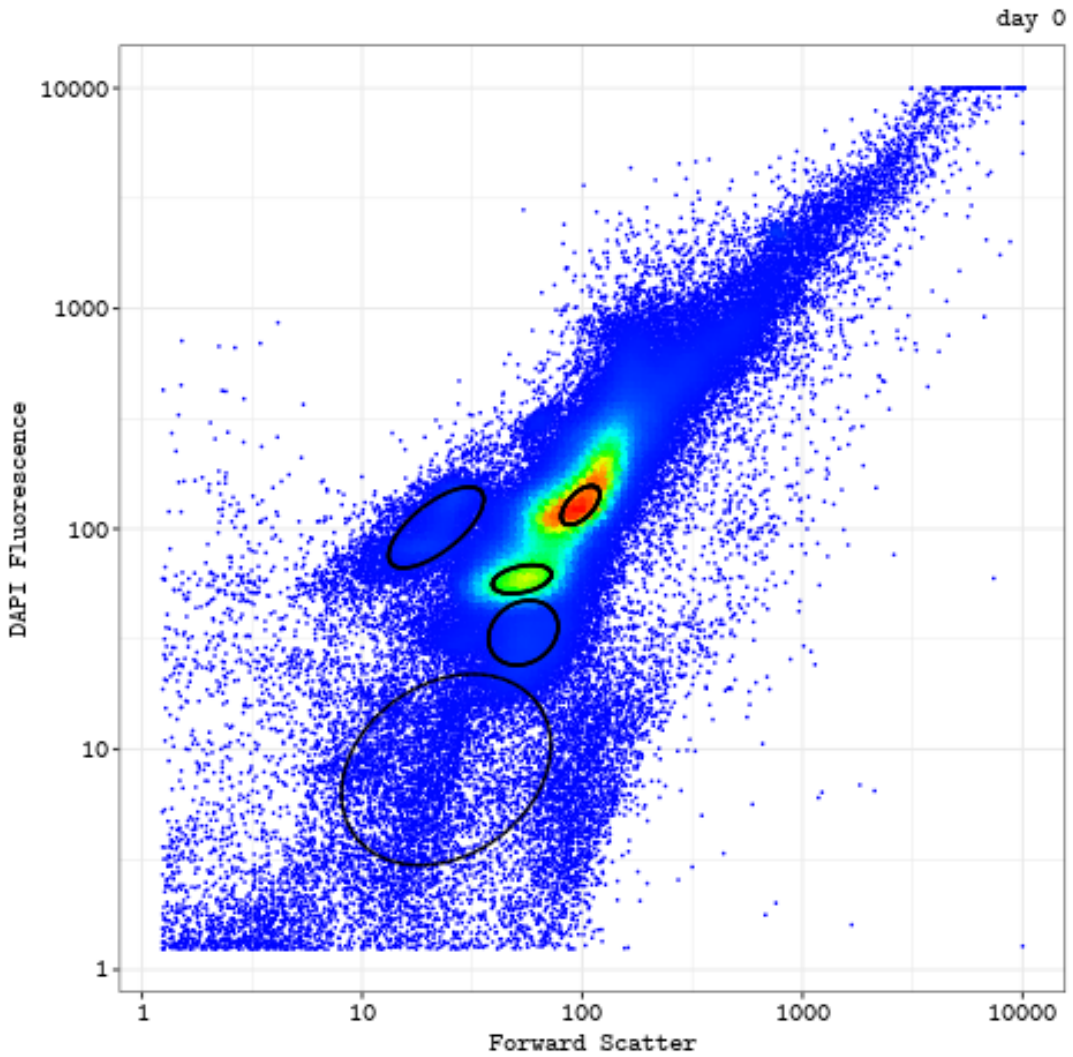


Example



Example 2

Automated gating of single samples



Gate Templates

- If we are given a time series of flow cytometry data, we would like to have a gate template that approximately fits every time point.

Gate Templates

- If we are given a time series of flow cytometry data, we would like to have a gate template that approximately fits every time point.
- Some ellipses are allowed to be empty if the corresponding sub-community is not present in that sample.

Gate Templates

- If we are given a time series of flow cytometry data, we would like to have a gate template that approximately fits every time point.
- Some ellipses are allowed to be empty if the corresponding sub-community is not present in that sample.
- Input: The ellipses of all the samples (combined into a single set)

Gate Templates

- If we are given a time series of flow cytometry data, we would like to have a gate template that approximately fits every time point.
- Some ellipses are allowed to be empty if the corresponding sub-community is not present in that sample.
- Input: The ellipses of all the samples (combined into a single set)
- Take out ellipses with a weight below a given threshold

Gate Templates

- If we are given a time series of flow cytometry data, we would like to have a gate template that approximately fits every time point.
- Some ellipses are allowed to be empty if the corresponding sub-community is not present in that sample.
- Input: The ellipses of all the samples (combined into a single set)
- Take out ellipses with a weight below a given threshold
- Remove the overlaps in this set of ellipses

Gate Templates

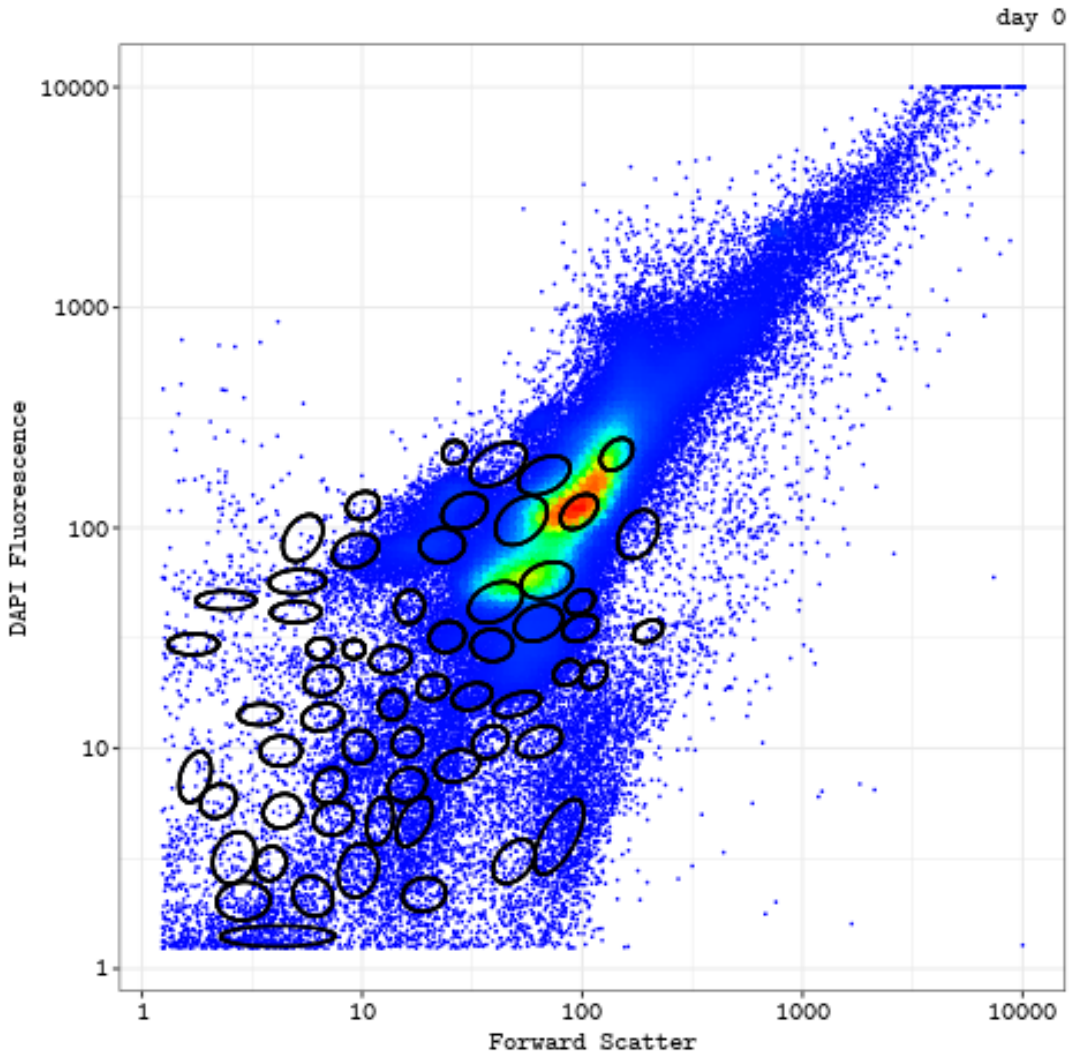
- If we are given a time series of flow cytometry data, we would like to have a gate template that approximately fits every time point.
- Some ellipses are allowed to be empty if the corresponding sub-community is not present in that sample.
- Input: The ellipses of all the samples (combined into a single set)
- Take out ellipses with a weight below a given threshold
- Remove the overlaps in this set of ellipses
- For computational reasons this is done recursively

Gate Templates

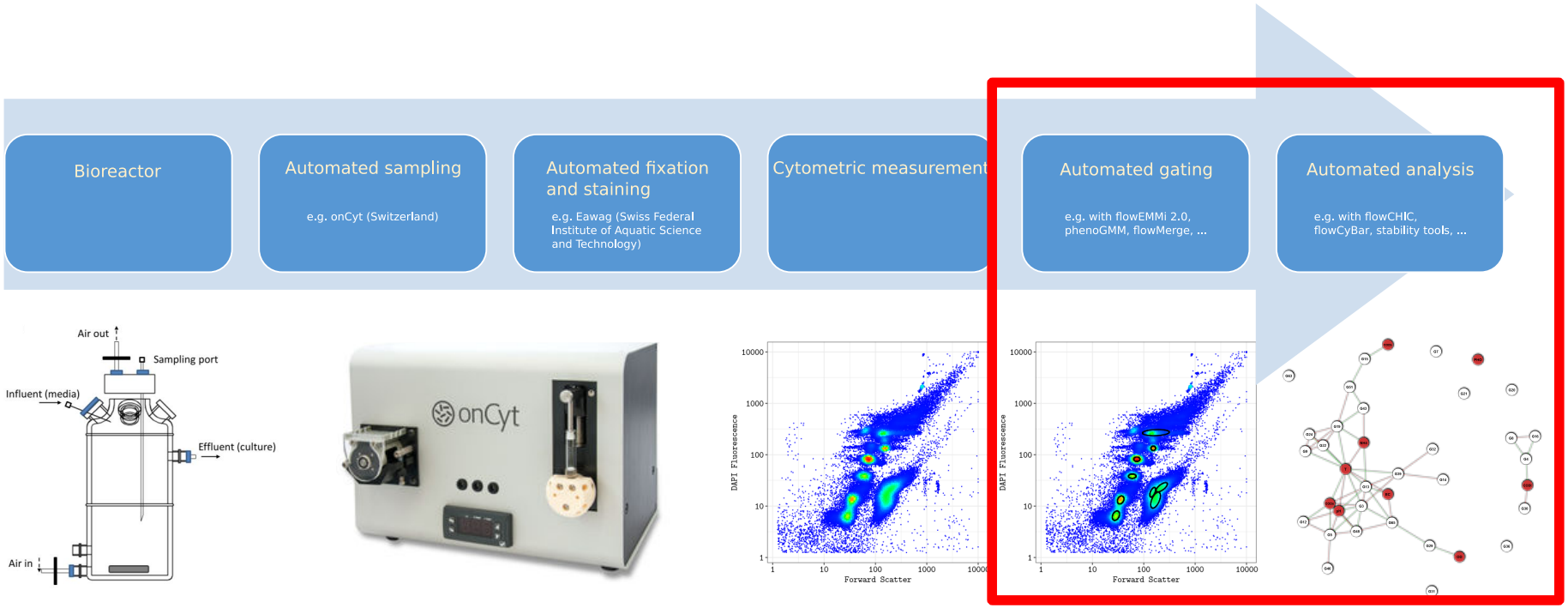
- If we are given a time series of flow cytometry data, we would like to have a gate template that approximately fits every time point.
- Some ellipses are allowed to be empty if the corresponding sub-community is not present in that sample.
- Input: The ellipses of all the samples (combined into a single set)
- Take out ellipses with a weight below a given threshold
- Remove the overlaps in this set of ellipses
- For computational reasons this is done recursively
- With the gate template we can calculate *cell numbers per gate and sample*

Example 3

Automatically generated gate template



Summary/Outlook



Liu et al., **Neutral mechanisms and niche differentiation in steady-state insular microbial communities revealed by single cell analysis**, *Environmental Microbiology, 2019*

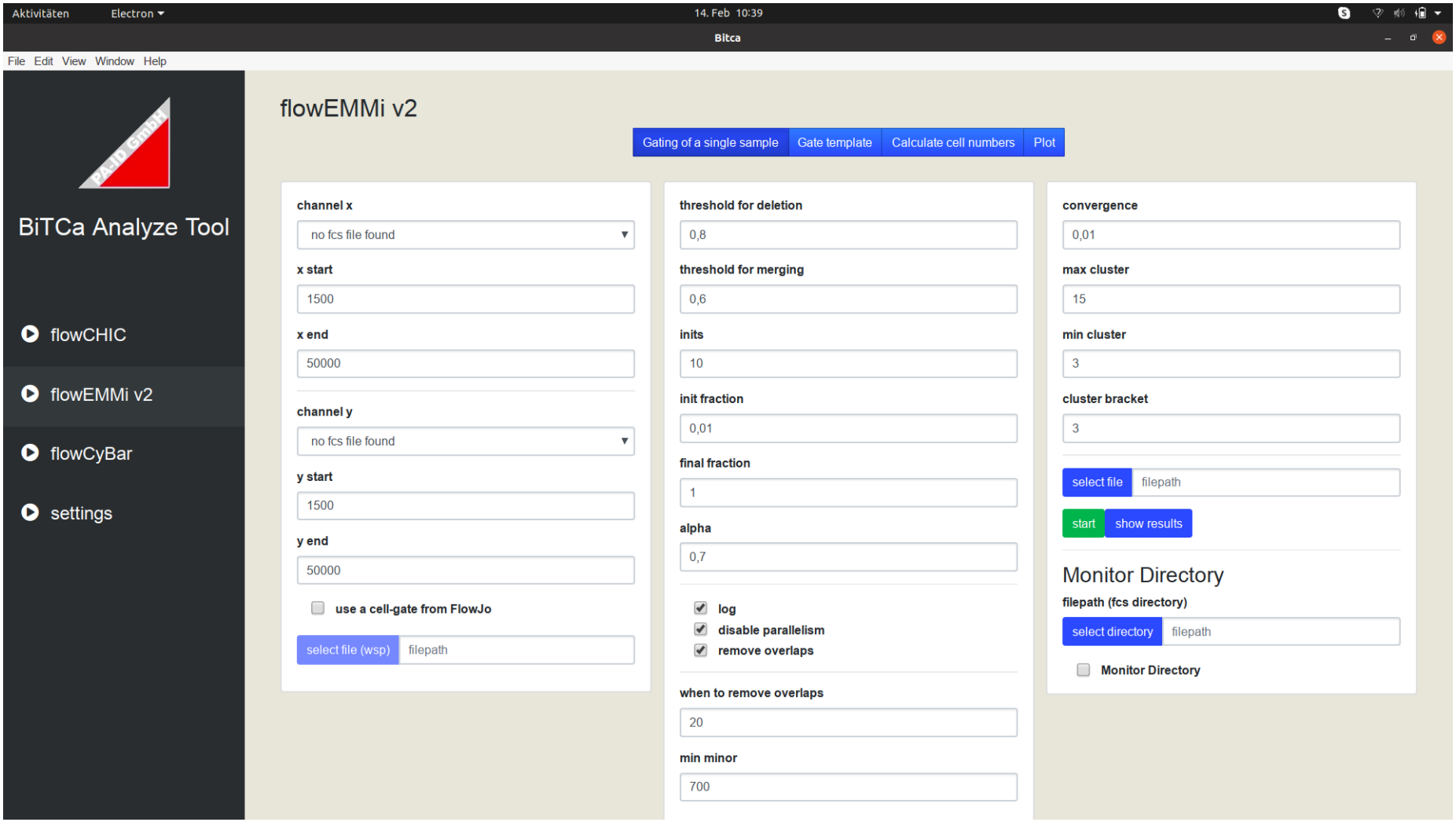
onCyt, *https://www.eawag.ch/de/news-agenda/news-plattform/newsarchiv/archiv-detail/online-durchflusszytometrie-sensor-fuer-bakterienkonzentrationen/*

Cichocki et al., **Bacterial mock communities as standards for reproducible cytometric microbiome analysis**, *Nature Protocols, 2020*

Ludwig et al., **flowEMMi: an automated model-based clustering tool for microbial cytometric data**, *BMC Bioinformatics, 2019*

Koch et al., **flowCyBar: Analyze flow cytometric data using gate information**, *R package version 1.30.0, 2019*

Summary/Outlook



Acknowledgements

UFZ HELMHOLTZ Zentrum für Umweltforschung

PA-ID GmbH

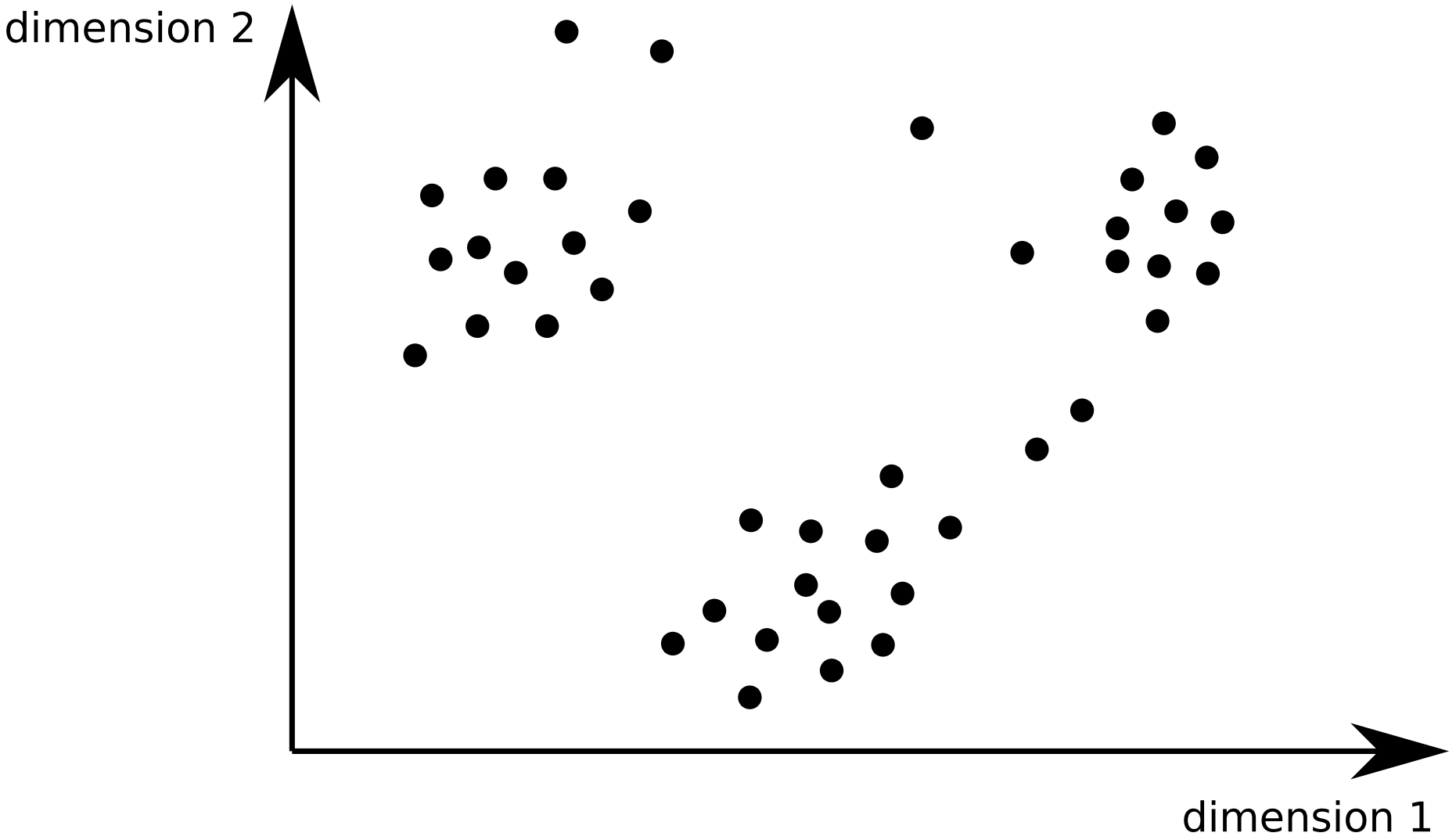
Bundesministerium für Wirtschaft und Energie

Bier Informatik est. 2002

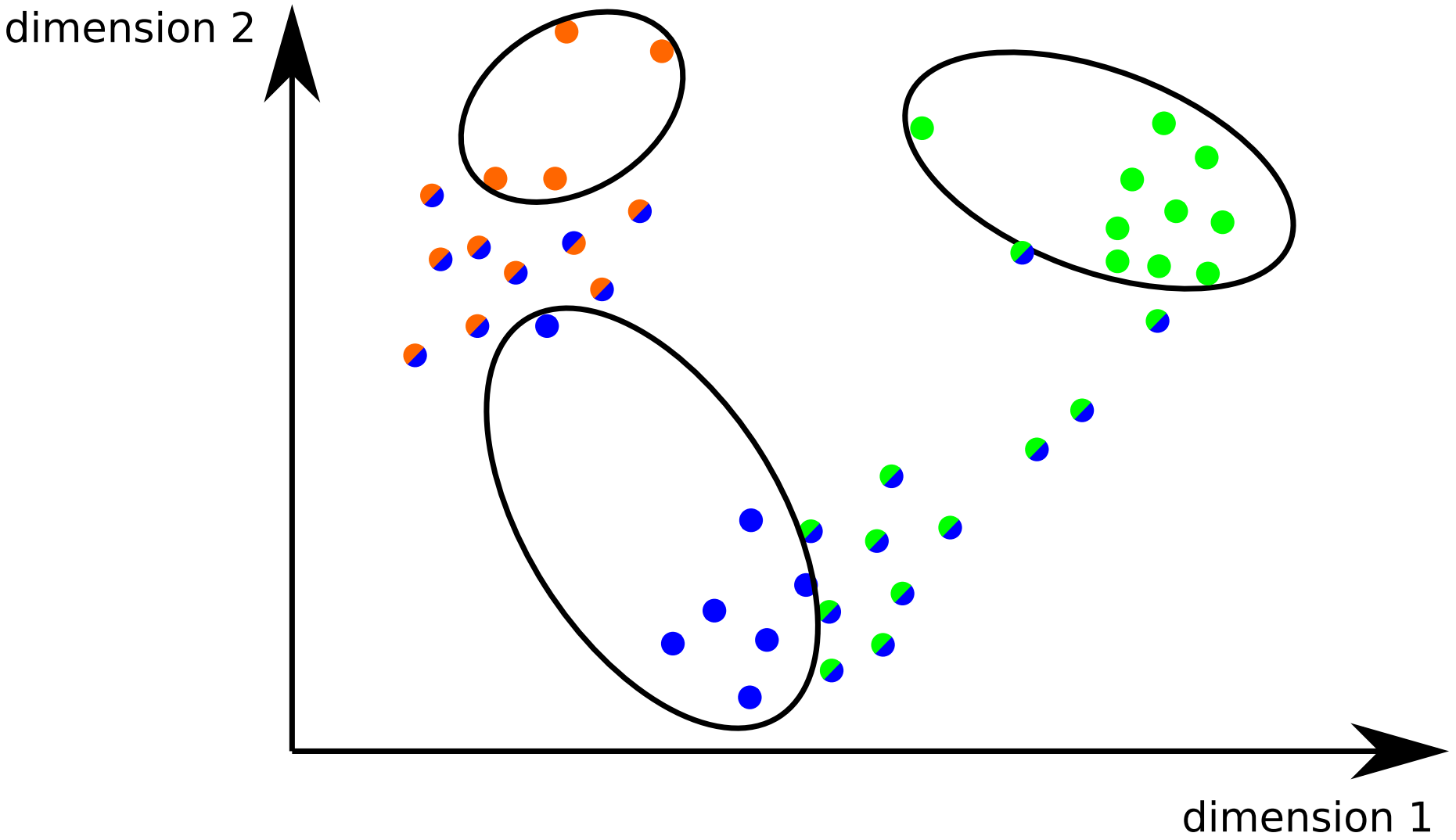
ScaDS.AI DRESDEN LEIPZIG

Thank you for your attention and your support!

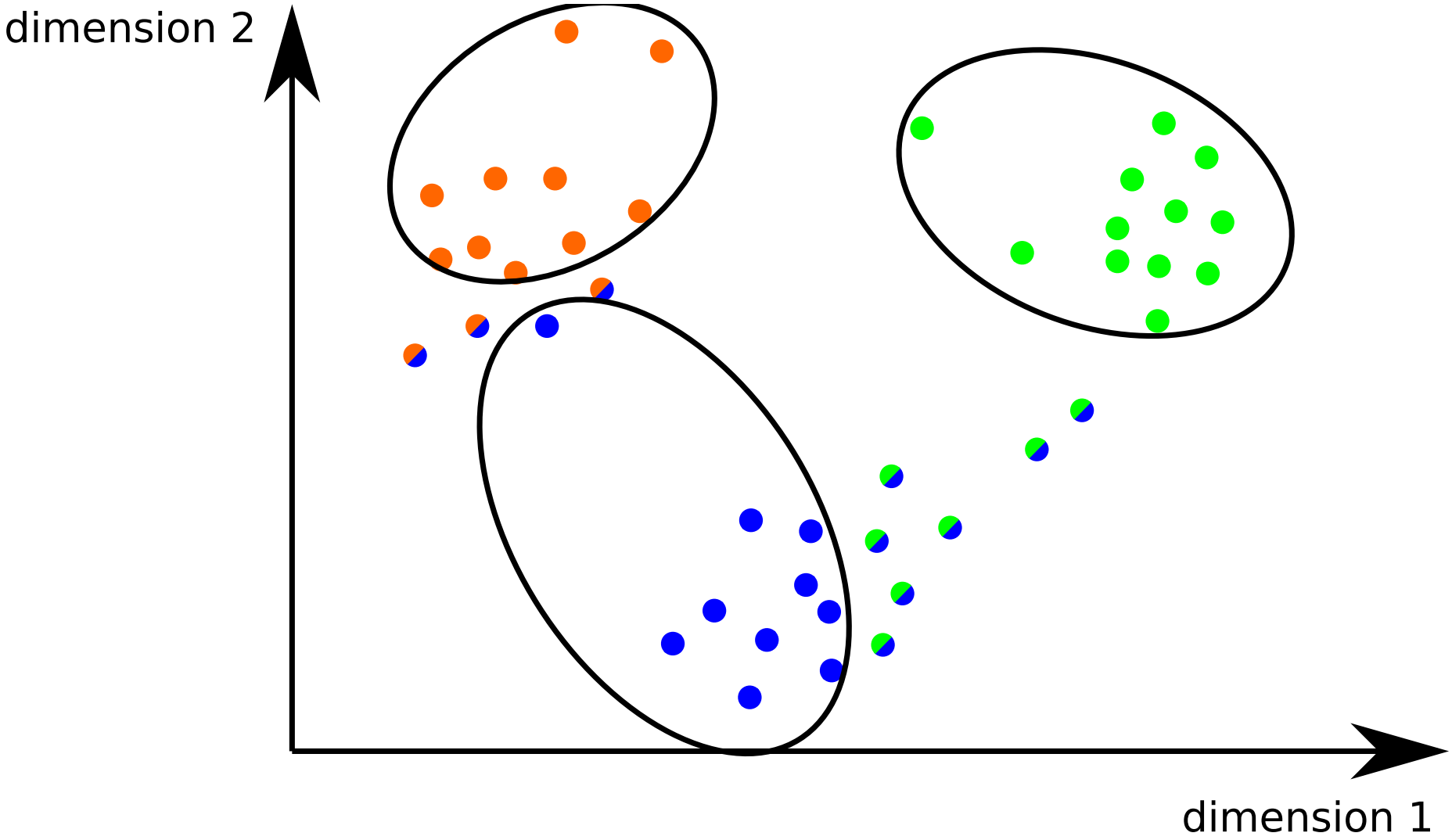
Appendix: The EM-algorithm



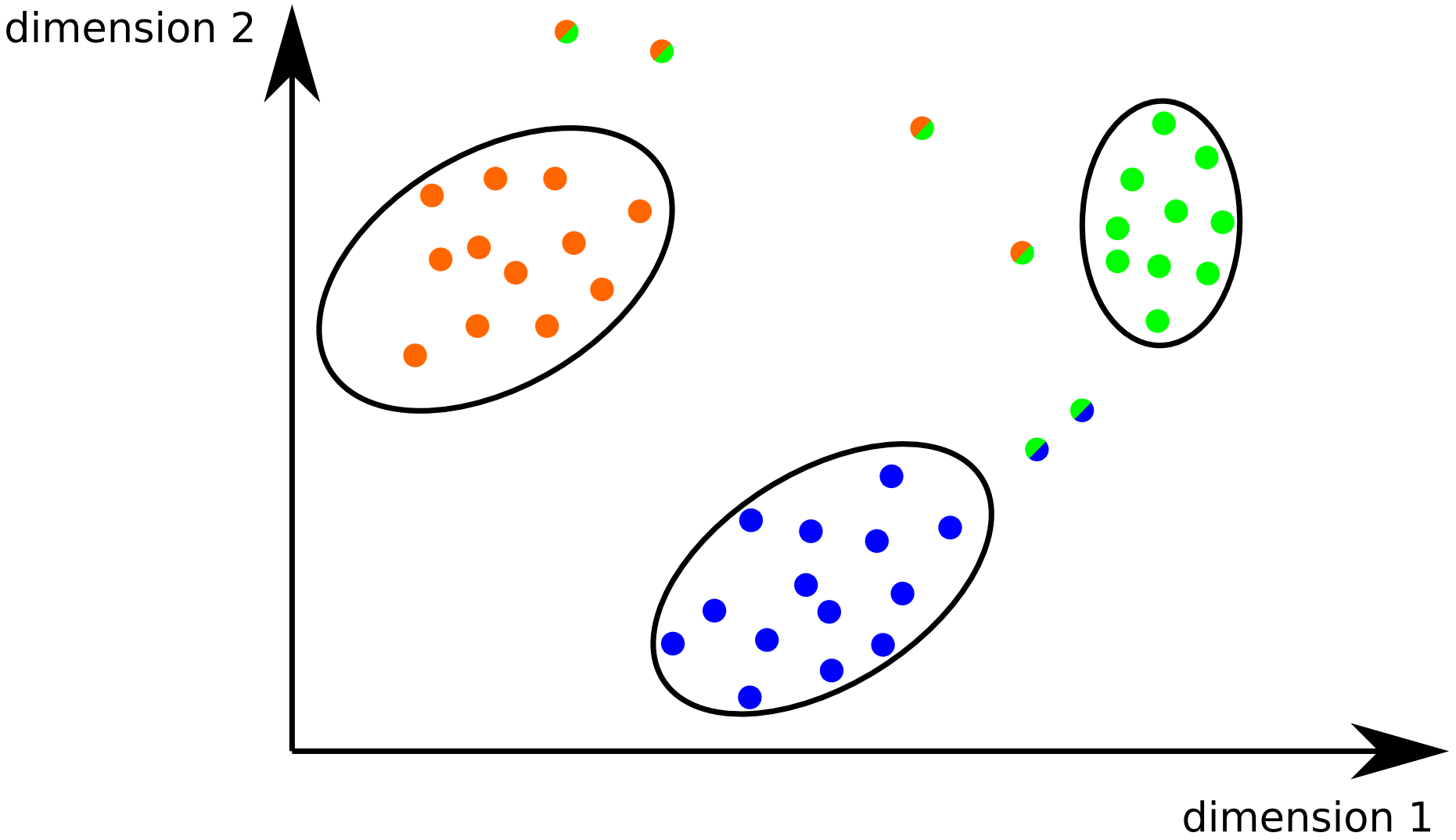
Appendix: The EM-algorithm



Appendix: The EM-algorithm

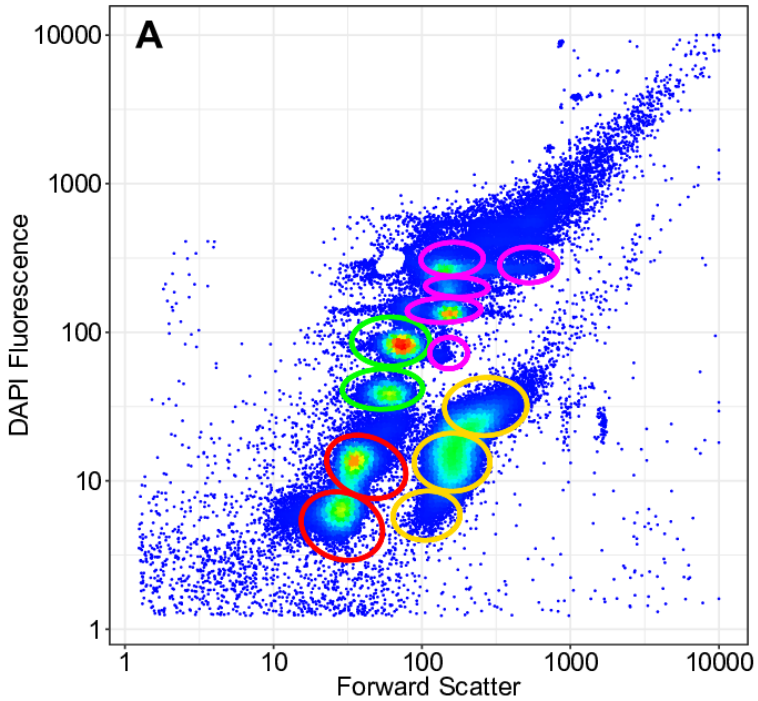


Appendix: The EM-algorithm

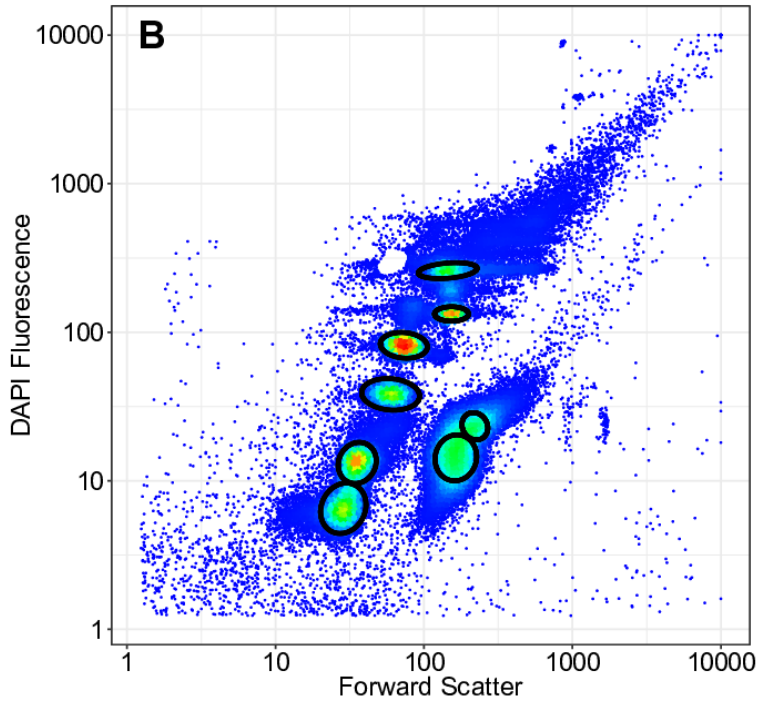


mock communities

manual gating



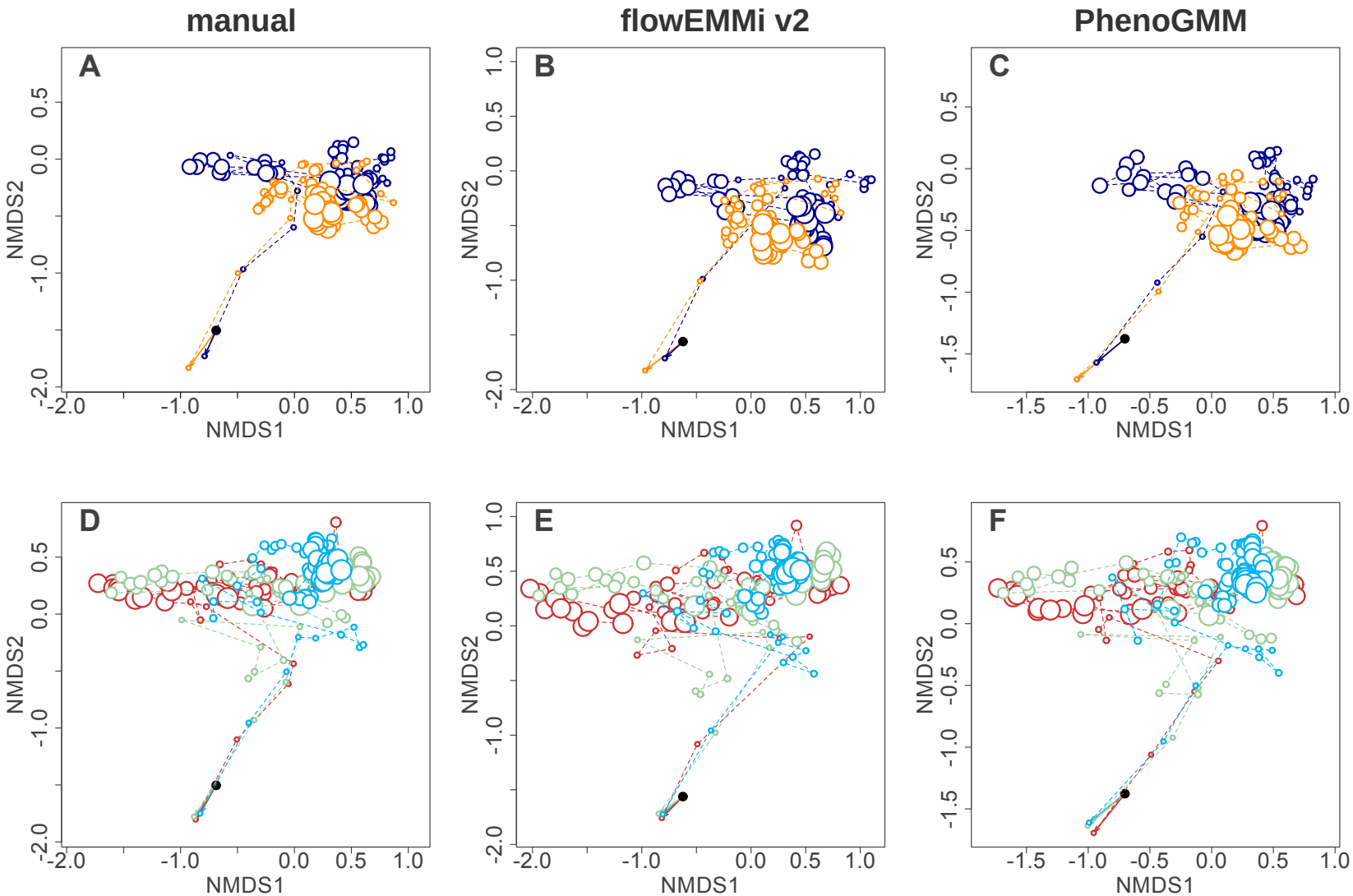
flowEMMi v2



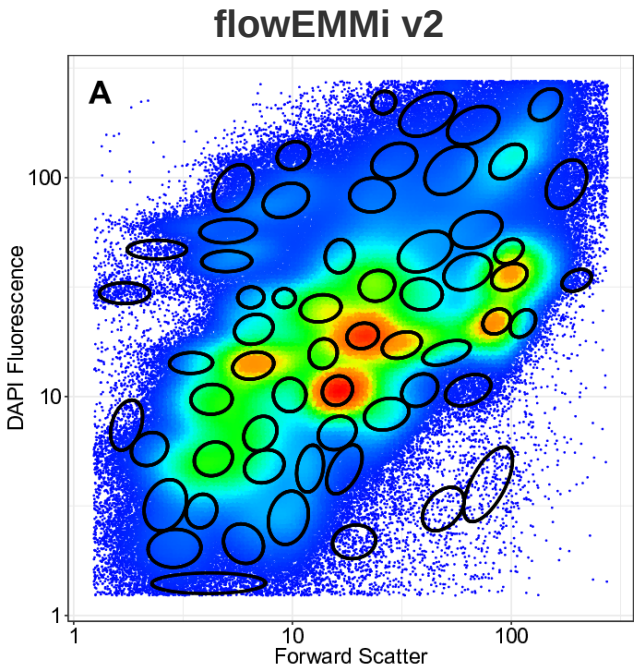
- *Stenotrophomonas rhizophila*
- *Escherichia coli*
- *Kocuria rhizophila*
- *Paenibacillus polymyxa*

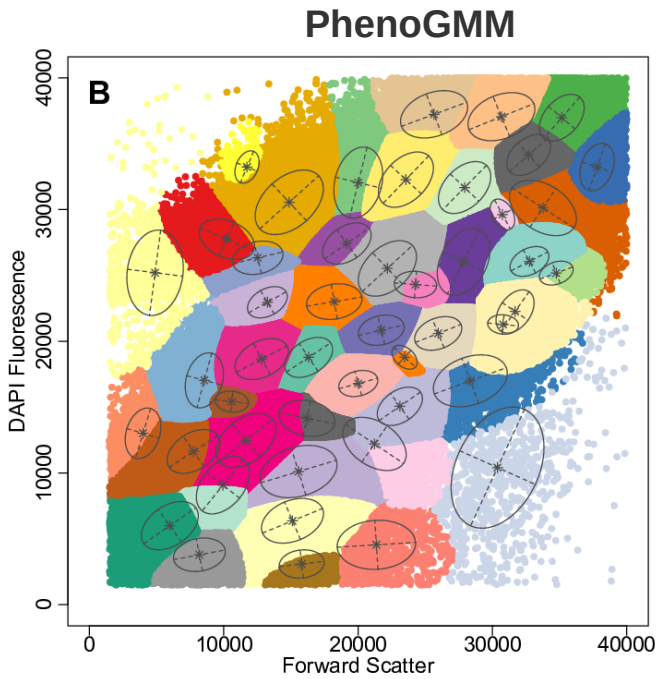
Cichocki et al., **Bacterial mock communities as standards for reproducible cytometric microbiome analysis**, *Nature Protocols, 2013*

Appendix: Community dynamics

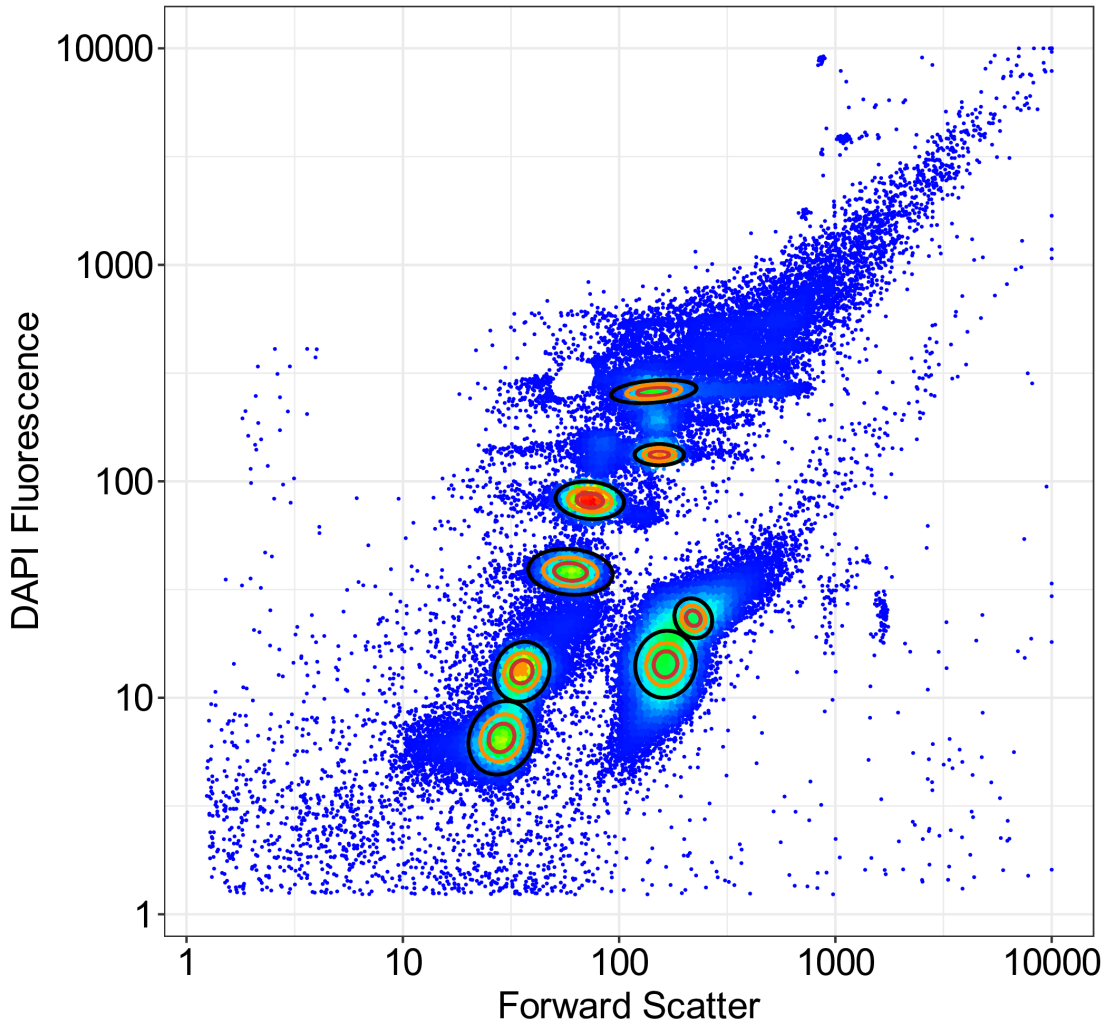


Appendix: Community dynamics

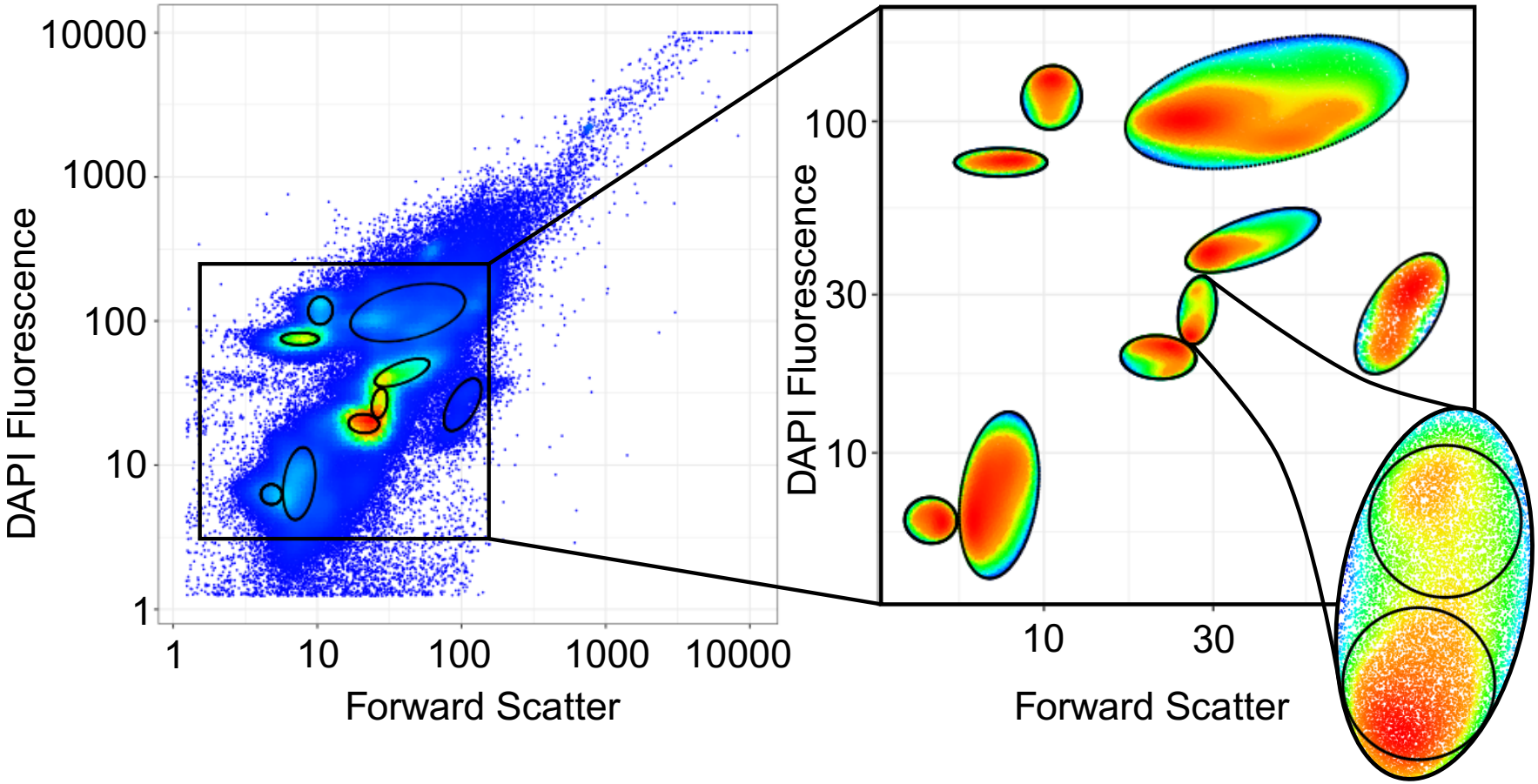




Appendix: Ellipse sizes



Appendix: Hierarchical Gating



Merging

Merging: The mathematical details

- Σ_1 and Σ_2 of E_1 and E_2 are both symmetric and positive definite $\Rightarrow$ use the Cholesky decomposition and re-scale them:

$$\Sigma_i = L_i \cdot D_i \cdot L_i^T \quad \text{for } i = 1, 2 \tag{1}$$

$$\Sigma_i' = L_i \cdot \frac{1}{w_i} \cdot D_i \cdot L_i^T \quad \text{for } i = 1, 2 \tag{2}$$

- Then, the new μ_c and Σ_c' of E_c', which will be re-scaled to E_c later, can be derived as the product of two Gaussians:

$$\mu_c = \left(\Sigma_1'^{-1} + \Sigma_2'^{-1}\right) \cdot \left(\Sigma_1'^{-1} \cdot \mu_1 + \Sigma_2'^{-1} \cdot \mu_2\right) \tag{3}$$

$$\Sigma_c' = \left(\Sigma_1'^{-1} + \Sigma_2'^{-1}\right)^{-1} = \Sigma_1' \cdot \left(\Sigma_1' + \Sigma_2'\right)^{-1} \cdot \Sigma_2' \tag{4}$$

- For the re-adjustment of the weights, we again need to decompose Σ_c' and multiply the diagonal elements by the sum of the weights.

$$\Sigma_c' = L_c' \cdot D_c' \cdot L_c'^T \tag{5}$$

$$\Sigma_c = L_c' \cdot (w_1 + w_2) \cdot D_c' \cdot L_c'^T \tag{6}$$