MONASH University

Clustering and visualisation tools to study high dimensional parameter spaces: B anomalies example

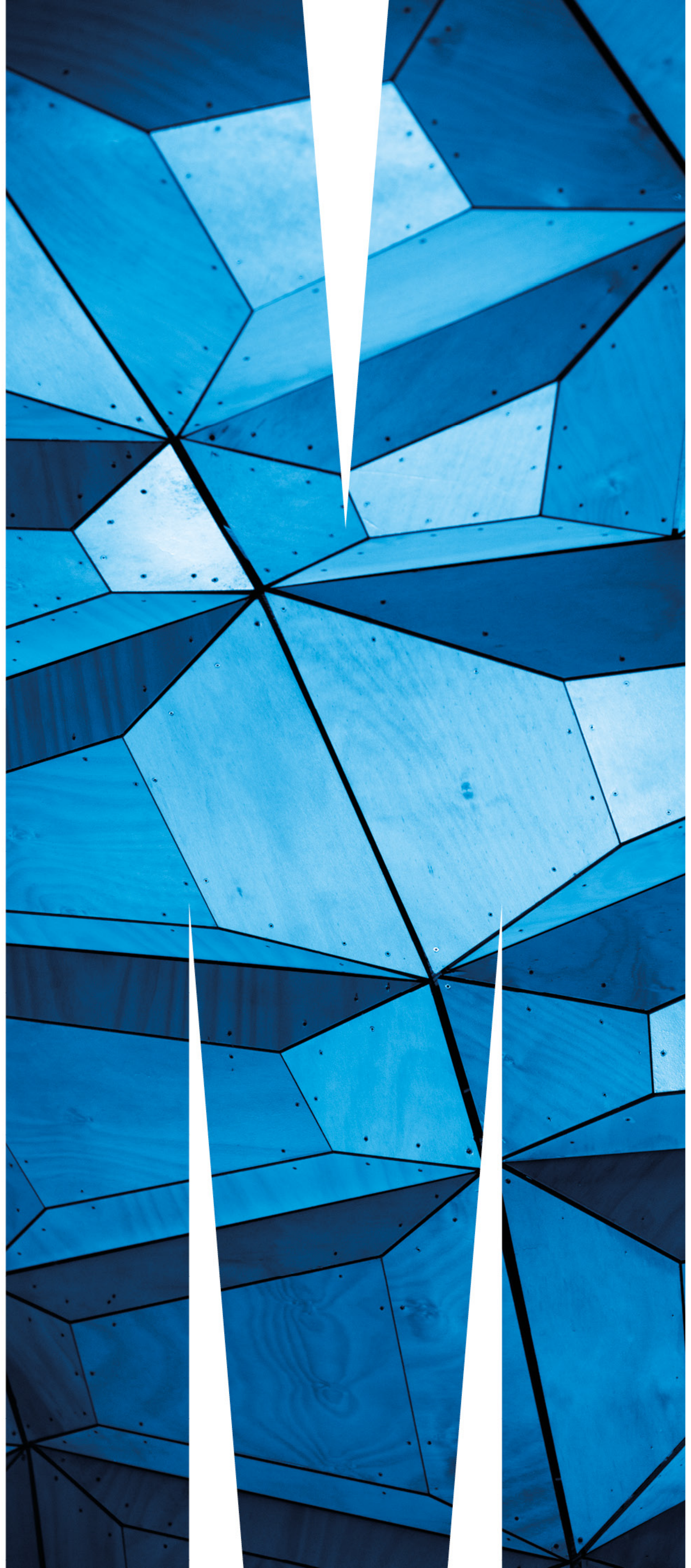
Corfu Summer Institute:
Workshop on the Standard Model and Beyond, 2022

German Valencia

based on work with Ursula Laa, D. Cook, and A. Aumann

Eur.Phys.J.Plus 137 (2022) 1, 145; Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2022.2035230 (2022);Journal of Computational and Graphical Statistics, 29:3, 681-687, DOI: 10.1080/10618600.2020.1777140

GROUP OF EIGHT AUSTRALIA



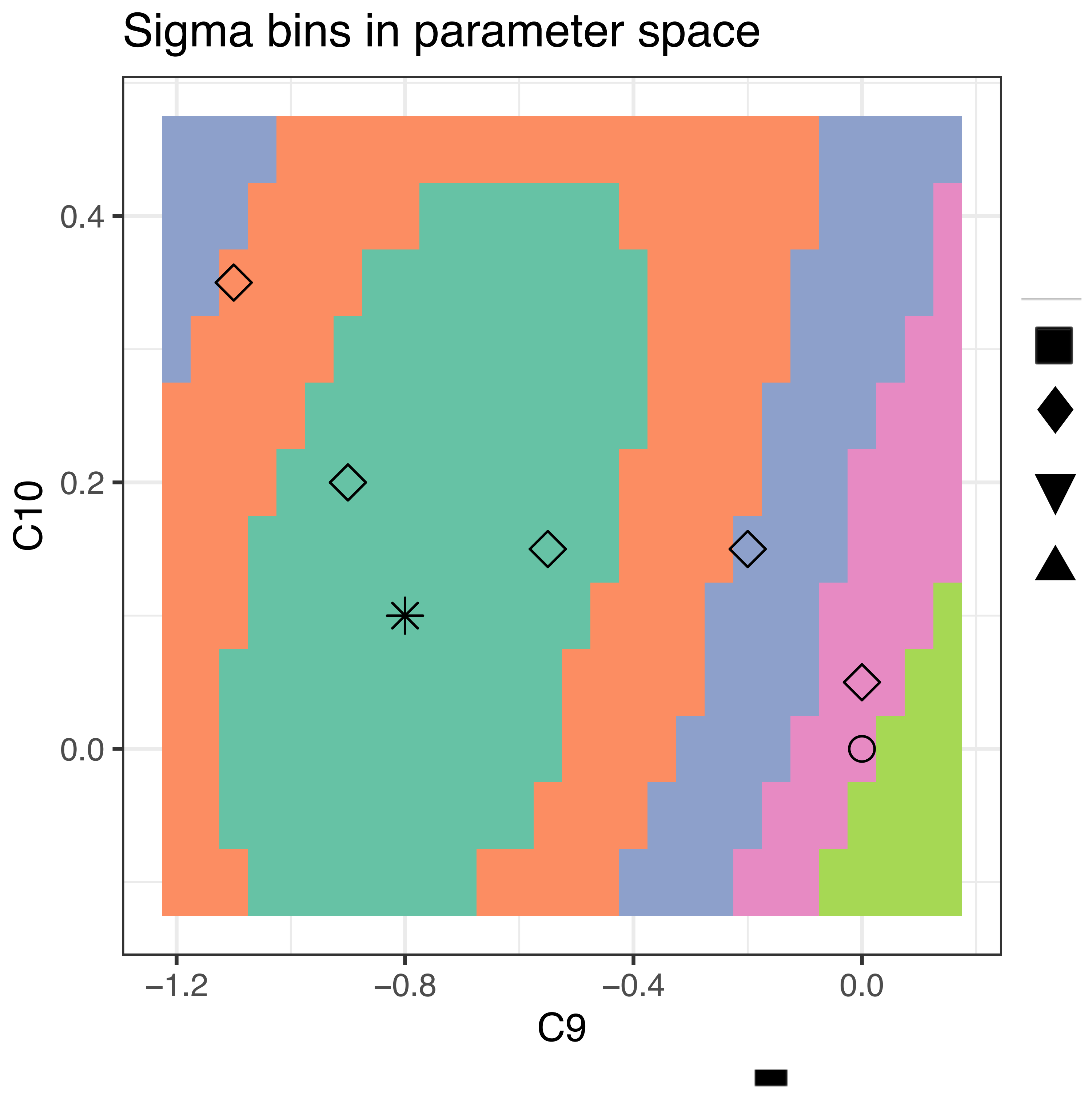
Two ways to see this talk

- If you work on B anomalies, we show different ways to look at the problem and gain some insight into it
- If you work on a different field, you have probably encounter problems that have many parameters or many observables where some of these tools may be useful

- The talk contains animations not visible in the pdf file.
  - Some can be generated by running the Shiny app https://github.com/uschiLaa/pandemonium
  - For others, contact one of us German.Valencia@monash.edu or ursula.laa@boku.ac.at

neutral B anomalies and global fits

- as we know, multiple observables in decays originating from the quark level transition $b \to s\ell^+\ell^-$ show deviations from the SM
- when these are quantified through global fits to hundreds of observables a deviation from the SM at the $\gtrsim 5\sigma$ level is found
- results that can be obtained from global fits include:
  - goodness of fit
  - best fit parameters
  - confidence level intervals
- confidence level intervals are a form of clustering
  - partitioning the parameter space based on $\Delta\chi^2$
  - everything determined by a distance to a reference point



6d-visualisation using grand tour

beyond a global fit

- clustering
  - partition of parameter space into “clusters” using all inter-point distances
    - more than one way to do it emphasising different aspects
    - the number of different groups is the resolving power of a specific data set
  - each cluster has a representative centroid
    - these provide a small set of benchmark points for detailed studies
  - isolate trends and effects of subsets of observables
- visualisation
  - visualise the collective dependence of the observables on the parameters
  - graphic displays of the observable space (more than 3d)
  - highlight the relative importance of observables: prioritising for further study
  - visual assessment of impact of correlations, dominant observables, tensions in the fit and others

reduced dimensionality $b \to s\ell^+\ell^-$

- we first reduce the dimensionality of parameter and observable spaces for conceptual clarity and to simplify visualisation

- the global fits use the Wilson coefficients in an effective Hamiltonian as parameters

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^\star \sum_{i,\ell=\mu,e} C_{i\ell}(\mu)\mathcal{O}_{i\ell}(\mu)$$

$$\mathcal{O}_7 = \frac{e}{16\pi^2} m_b (\bar{s}\sigma_{\mu\nu} P_R b) F^{\mu\nu}, \qquad \mathcal{O}_{7'} = \frac{e}{16\pi^2} m_b (\bar{s}\sigma_{\mu\nu} P_L b) F^{\mu\nu},$$

$$\mathcal{O}_9 = \frac{e^2}{16\pi^2} (\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu \ell), \qquad \mathcal{O}_{9'} = \frac{e^2}{16\pi^2} (\bar{s}\gamma_\mu P_R b)(\bar{\ell}\gamma^\mu \ell),$$

$$\mathcal{O}_{10} = \frac{e^2}{16\pi^2} (\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu \gamma_5 \ell), \qquad \mathcal{O}_{10'} = \frac{e^2}{16\pi^2} (\bar{s}\gamma_\mu P_R b)(\bar{\ell}\gamma^\mu \gamma_5 \ell),$$

- assuming NP only for the muons and CP conservation
  - global fits show most important ones are C_9^μ, C_{10}^μ — two parameter study
  - illustrate case with additional $C_{9'}^\mu, C_{10'}^\mu$ — four parameter study

neutral B anomalies: choose 14 observables

	ID	Observable	Exp.	ID in [4]
★	1	$P_5'(B \to K^* \mu\mu)[0.1 - 0.98]$	0.52 ± 0.10	20
	2	$P_5'(B \to K^* \mu\mu)[1.1 - 2.5]$	0.36 ± 0.12	28
	3	$P_5'(B \to K^* \mu\mu)[2.5 - 4]$	-0.15 ± 0.14	36
★	4	$P_5'(B \to K^* \mu\mu)[4 - 6]$	-0.39 ± 0.11	44
★	5	$P_5'(B \to K^* \mu\mu)[6 - 8]$	-0.58 ± 0.09	52
	6	$P_5'(B \to K^* \mu\mu)[15 - 19]$	-0.67 ± 0.06	60
	7	$P_2(B \to K^* \mu\mu)[0.1 - 0.98]$	0 ± 0.04	17
	8	$P_2(B \to K^* \mu\mu)[1.1 - 2.5]$	-0.44 ± 0.10	25
	9	$P_2(B \to K^* \mu\mu)[2.5 - 4]$	-0.19 ± 0.12	33
★	10	$P_2(B \to K^* \mu\mu)[4 - 6]$	0.10 ± 0.07	41
★	11	$P_2(B \to K^* \mu\mu)[6 - 8]$	0.21 ± 0.05	49
★	12	$P_2(B \to K^* \mu\mu)[15 - 19]$	0.36 ± 0.02	57
★	13	$R_K(B^+ \to K^+)[1.1 - 6]$	0.86 ± 0.06	98
★	14	$R_{K^*}(B^0 \to K^{0*})[1.1 - 6]$	0.73 ± 0.11	100

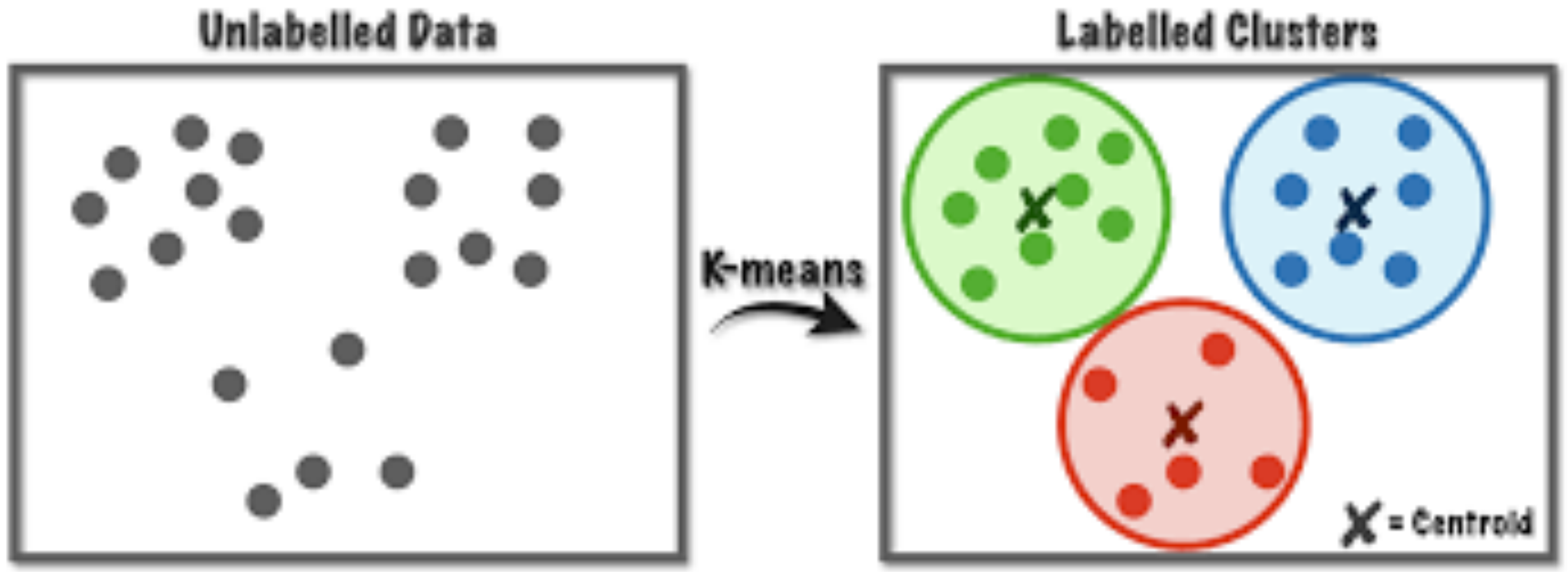
★ B. Capdevila, U. Laa, G.V. *Eur.Phys.J.C* 79 (2019) 6, 462

- select a subset of observables based on the ranking analysis of *Eur.Phys.J.C* 79 (2019) 6, 462 showing importance for global fits
- ★ singled out as most important for C_9^μ, C_{10}^μ
- complete P_5', P_2 bins
- ★ singled out as important for $C_{9'}^\mu, C_{10'}^\mu$
- best 2d fit from this set (flavio) lies 3.7σ from the SM at $(C_9, C_{10}) = (-0.8, 0.1)$

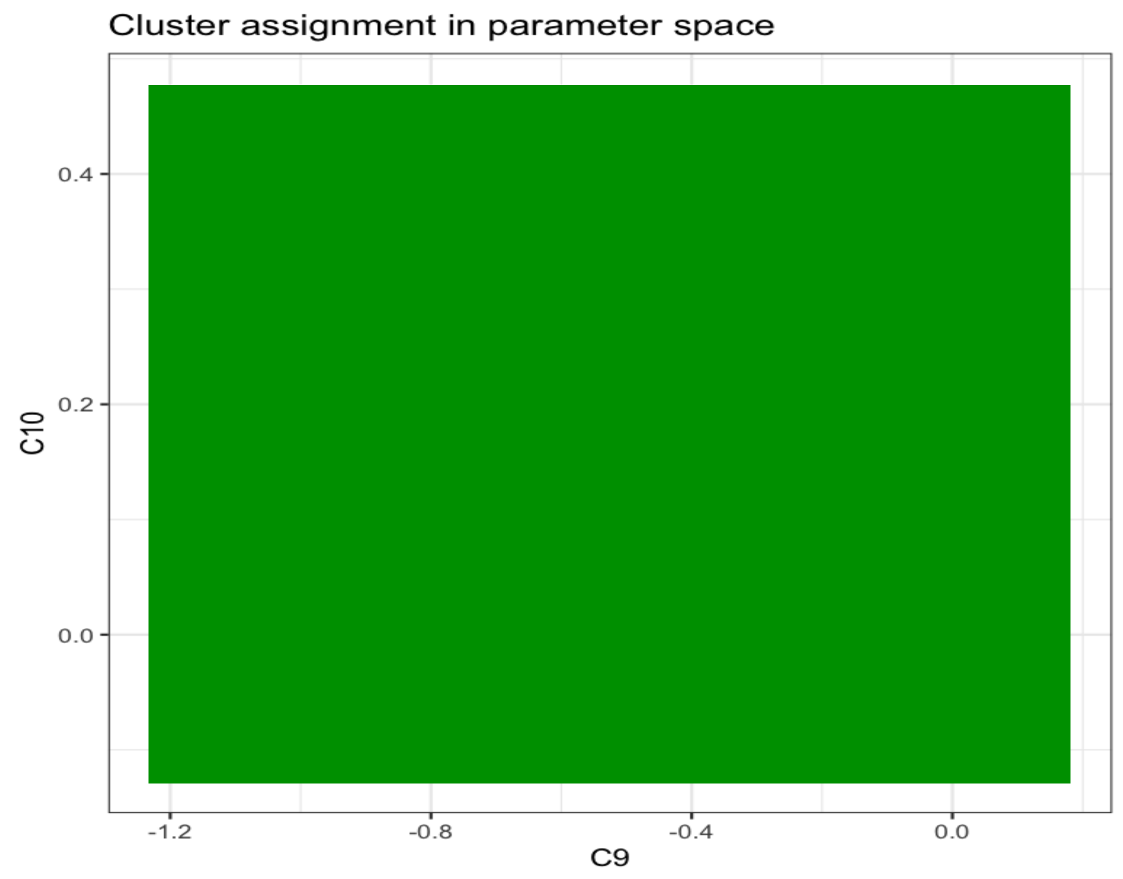
clustering: some definitions

clustering: what are we doing?

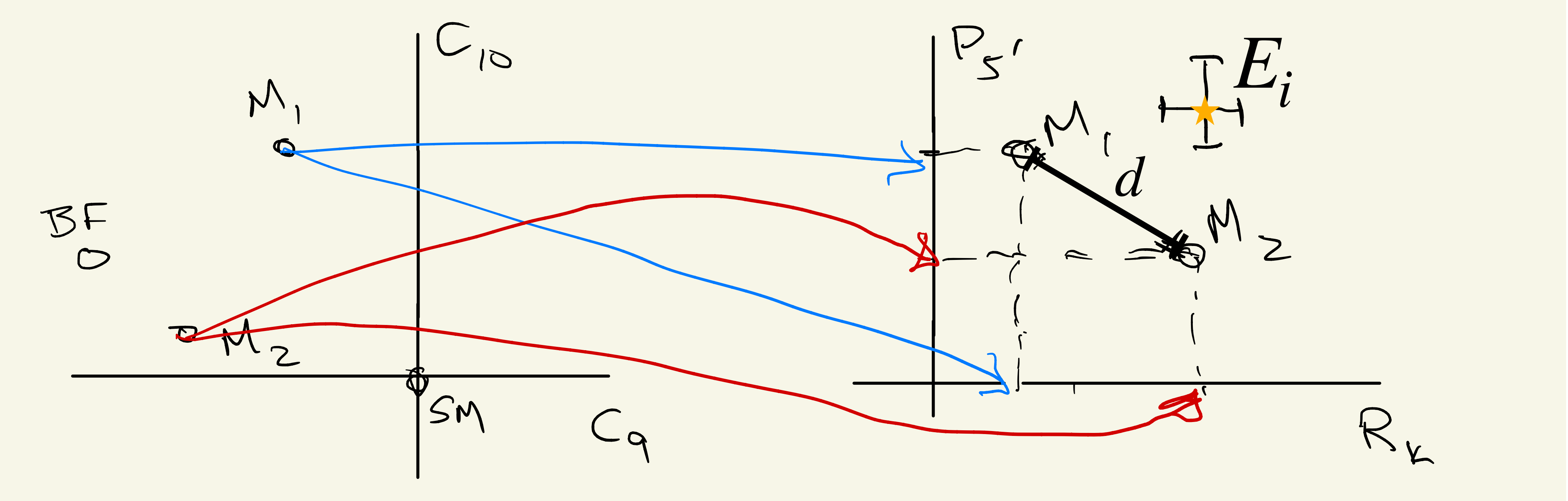
- Usually we have a picture like this for clustering



- How to partition parameter space that looks like this?

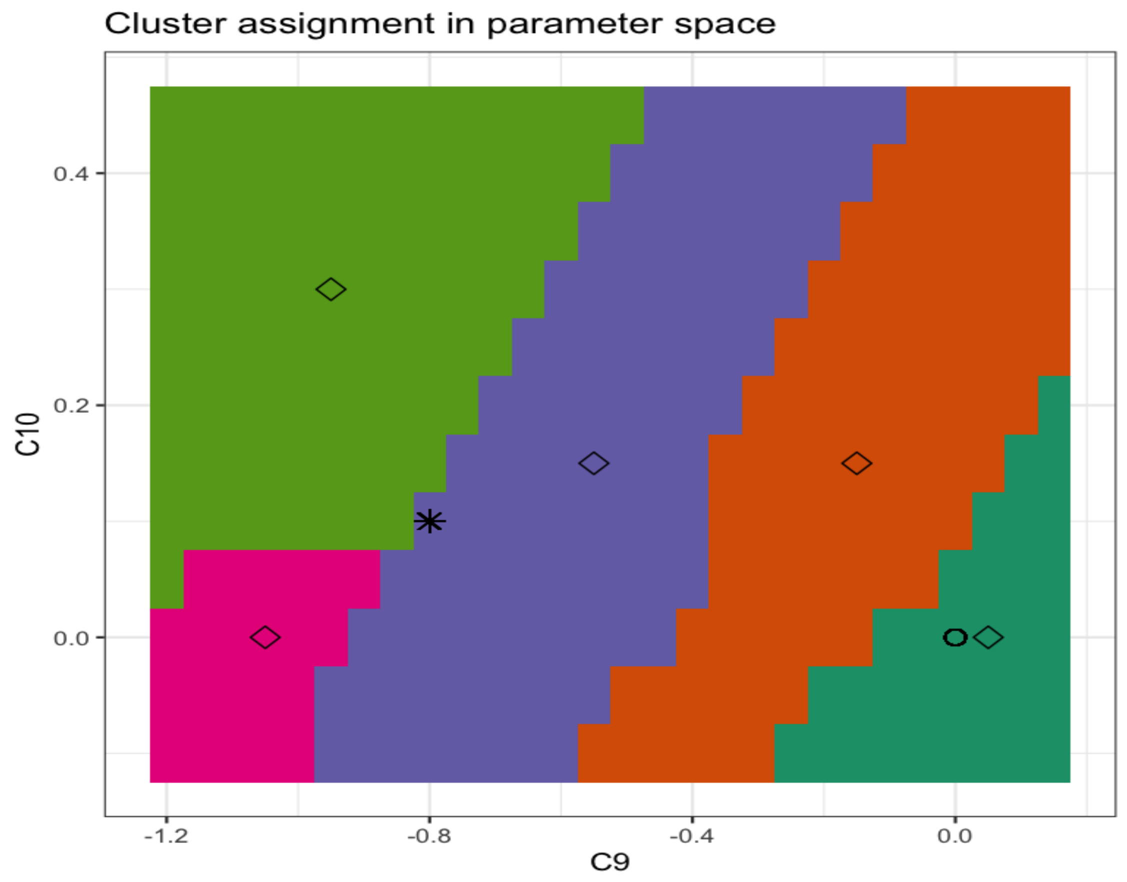


metric in observable space



parameter space observable space

clustering algorithm



distance between models

- coordinates: observable measured from a reference point in units of uncertainty: "pulls" (Σ = total covariance matrix)

$$Y_{ki} = \sum_j \Sigma_{ij}^{-1/2}(X_{kj} - O_j) \approx \sum_j \frac{1}{\sqrt{(\Sigma^{-1})_{ii}}}(\Sigma^{-1})_{ij}(X_{kj} - O_j)$$

- X_{ki} is the prediction of model k for observable i
- the "origin' O_i is any reference point: experiment E_i, SM, $\cdots$
- model k $\xrightarrow[d=2]{\text{parameter space}}$ (C_{9k}, C_{10k}) $\xrightarrow[d=i=14]{\text{observable space}}$ Y_{ki}

distance between models

- the χ^2 for a model is the squared Euclidean distance to experiment

$$\chi_k^2 = \sum_{i,j} [X_{ki} - E_i](\Sigma^{exp} + \Sigma^{th})^{-1}_{ij}[X_{kj} - E_j] = \sum_i Y_{ki}^2$$

- analogously, the (Mahalanobis) distance between models

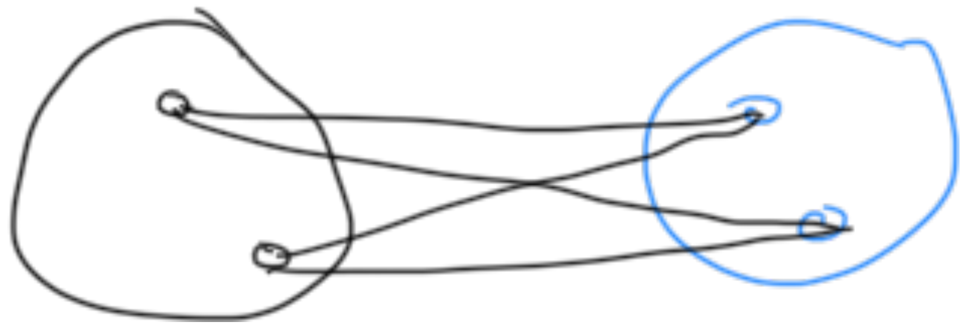
$$d_{\chi_k^2}(X_k, X_l) = \sum_{i,j} [X_{ki} - X_{li}](\Sigma^{exp} + \Sigma^{th})^{-1}_{ij}[X_{kj} - X_{lj}] = \sum_i (Y_{ki} - Y_{li})^2$$

- the last equality follows if Σ does not depend on the model
  - Results do not depend on the origin
- clustering takes into account all inter-point distances and not just distance to a reference point, can be done before you have measurements
- In this example, we use a distance that can be interpreted as a $\Delta\chi^2$

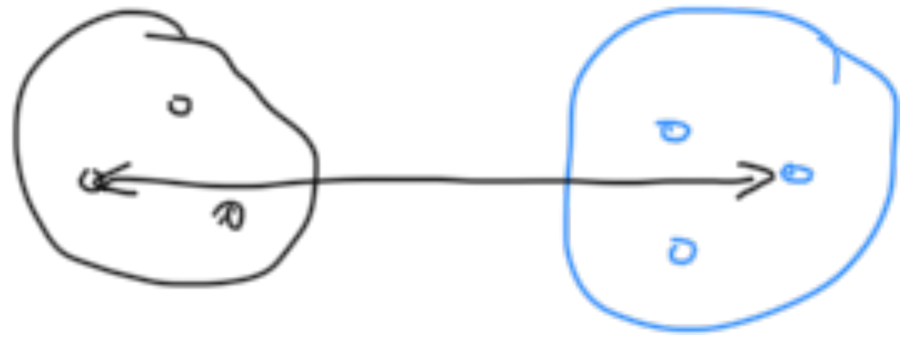
linkage: distance between clusters

previously used in particle physics applications

- average: the distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster

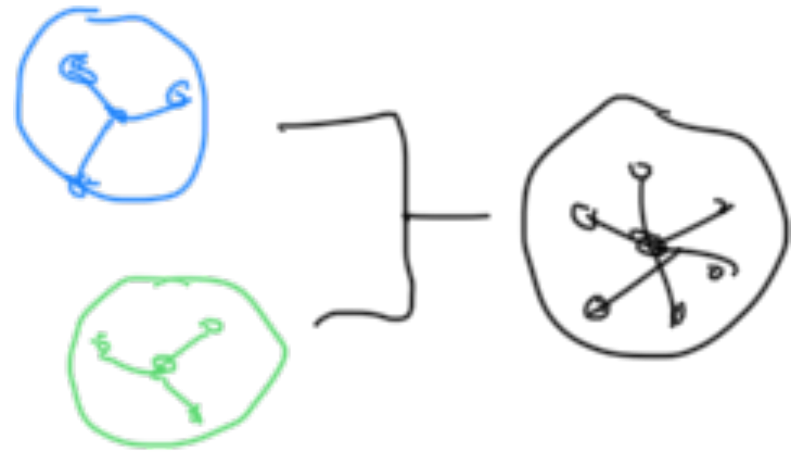


- complete: the distance between two clusters is defined as the longest distance between two points in each cluster



we will also use

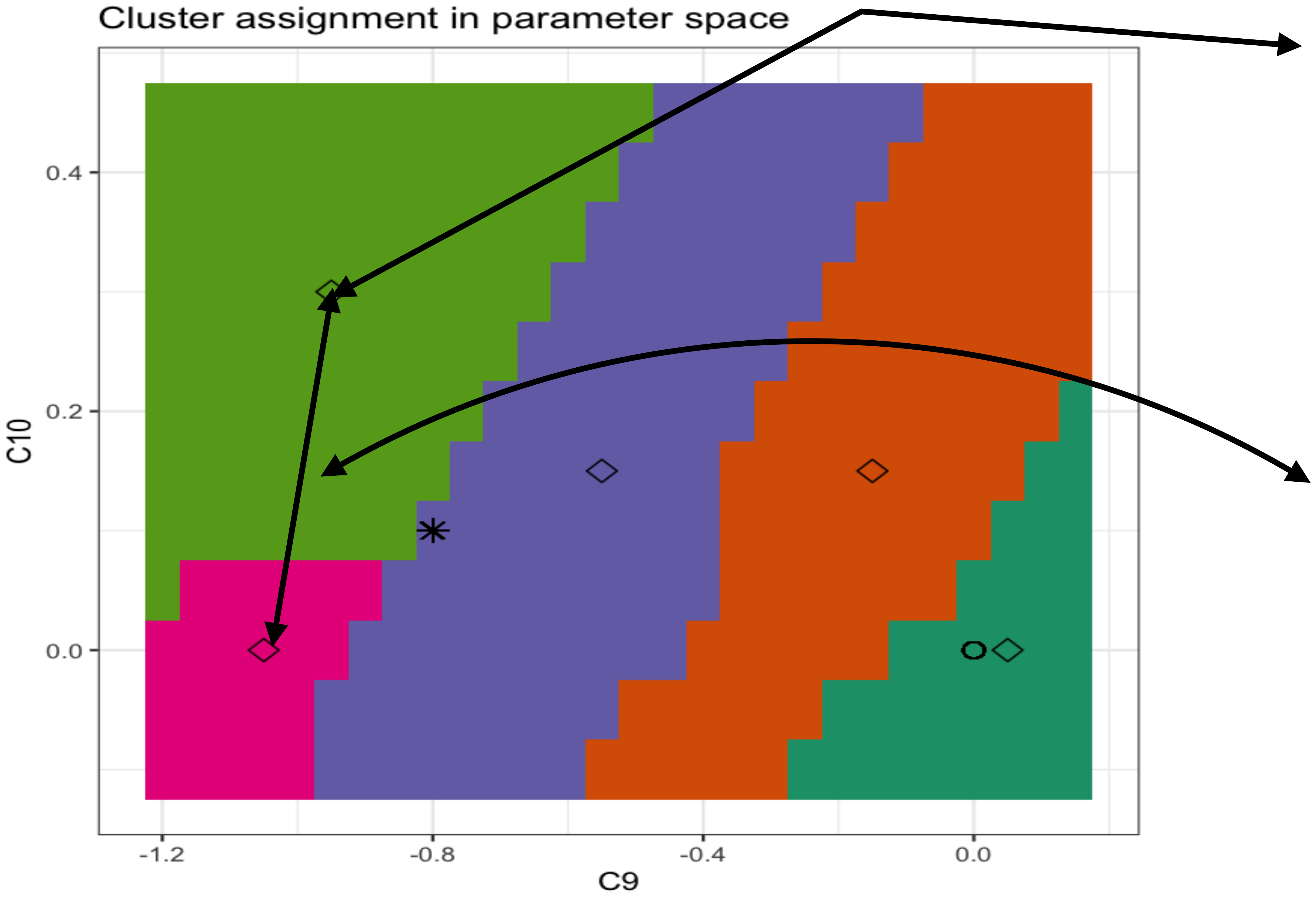
- ward: instead of measuring the distance directly, it analyzes the variance of clusters. Ward's method says that the distance between two clusters, A and B, is how much the sum of squares will increase when we merge them.
  - minimizes inter-cluster dissimilarity



centroid and “radius” (maximum within distance)

- want centroids to be representative of their cluster
- centroids serve as benchmark points for further study
- The centroid c_j is the member of the cluster C_j which minimises
$$f(c, C_j) = \sum_{x_i \in C_j} d(c, x_i)^2$$
- The “radius” of the cluster is
$$r_j = \max_{x_i \in C_j} d(c_j, x_i)$$

one possibility "one sigma" clusters



- using a suitable distance function
- if the BF to (future) experiments fell at this centroid, all the light green points would be in the 1σ region, $\Delta\chi^2 \leq 2.3$ for 2 parameters
- note that there is no one centroid that is singled out by a global fit. At this stage, there not need be a fit or even measurements
- the distance between any two centroids would be at least 1σ: $\Delta\chi^2 \geq 2.3$
- caveats:
- there will always be points as close to each other as we want that sit on different clusters
- boundaries will shift if the parameter range is changed

back to our application to B anomalies

cluster the 14 observables

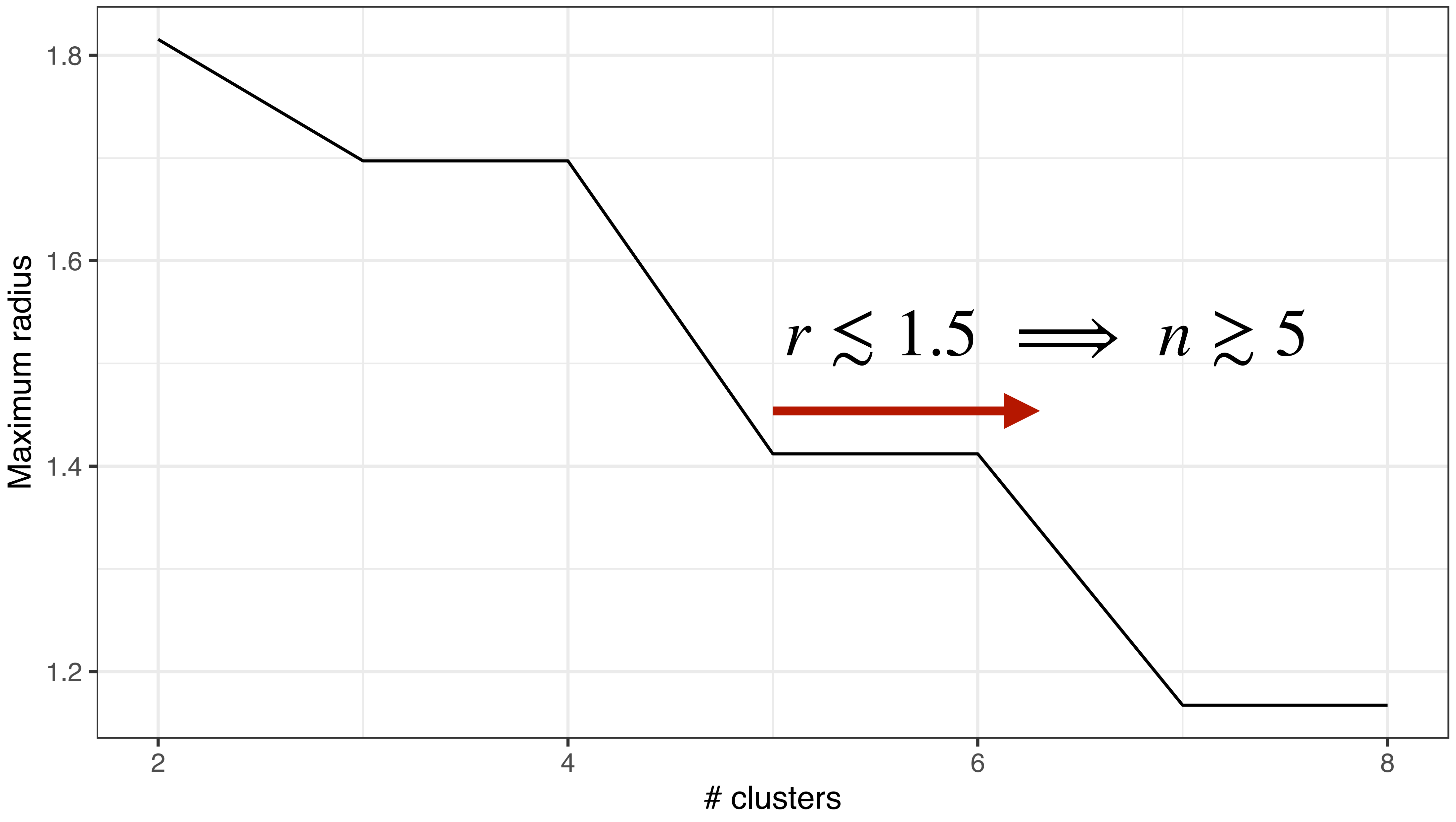
- generate models (sets of 14 predictions) on a uniform grid of parameters C_9^μ, C_{10}^μ
  - Uniform grid was needed for original Shiny app, not anymore
  - All data shown was generated with flavio (D. Straub, "flavio: arXiv:1810.08132)
- the grid includes the BF and the SM and points in the region between them
- in what follows we use either Euclidean distance with Ward or average linkage, or maximum distance with complete linkage

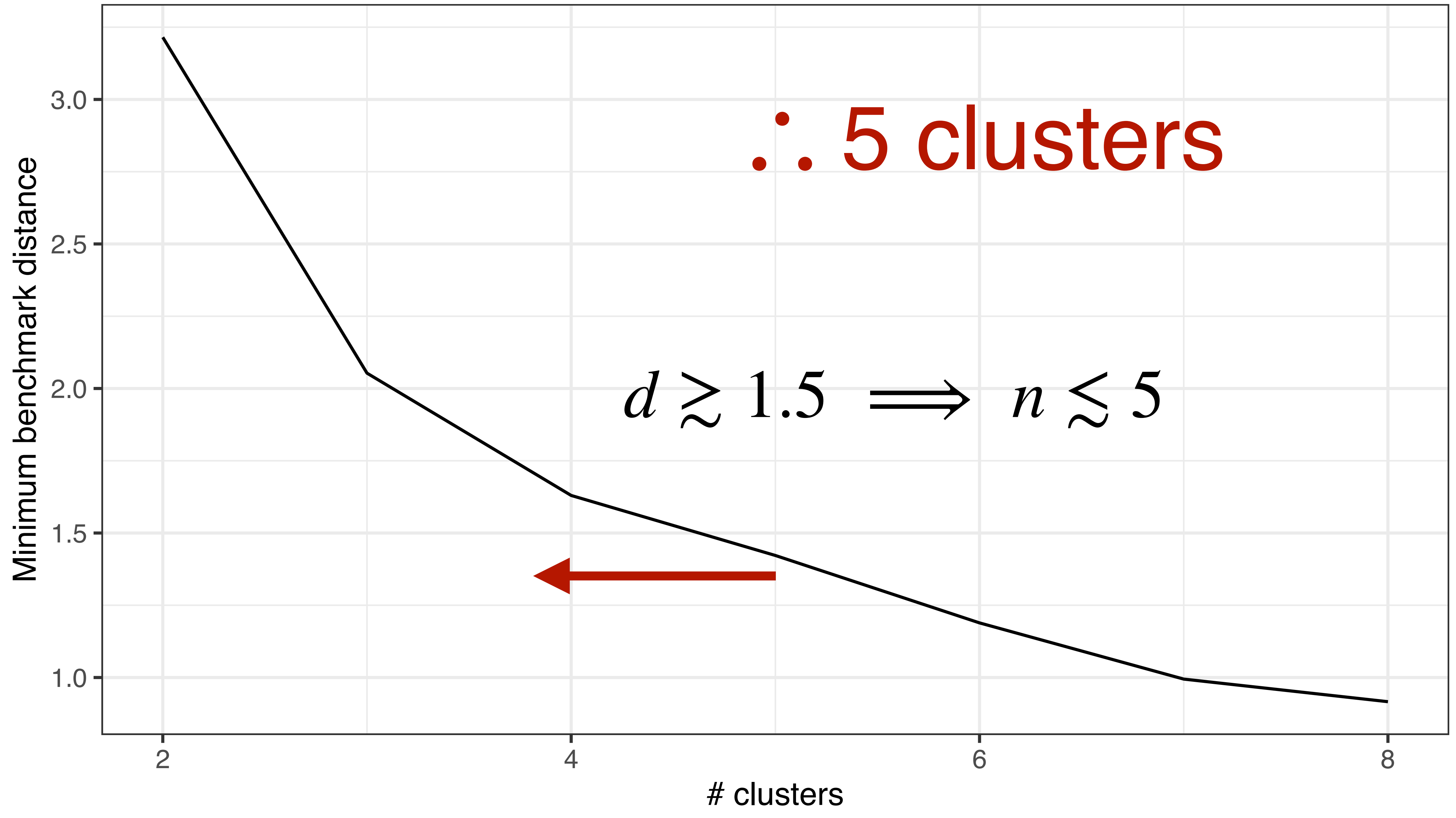
how many clusters? -euclidean distance

- square of euclidean distance on (“pulls”) $Y_k \approx \Delta\chi^2$

- cluster as set of points that are indistinguishable from each other at some level of confidence: fix the maximum radius

- separate clusters are different at some level of confidence: fix the minimum distance between cluster centroids

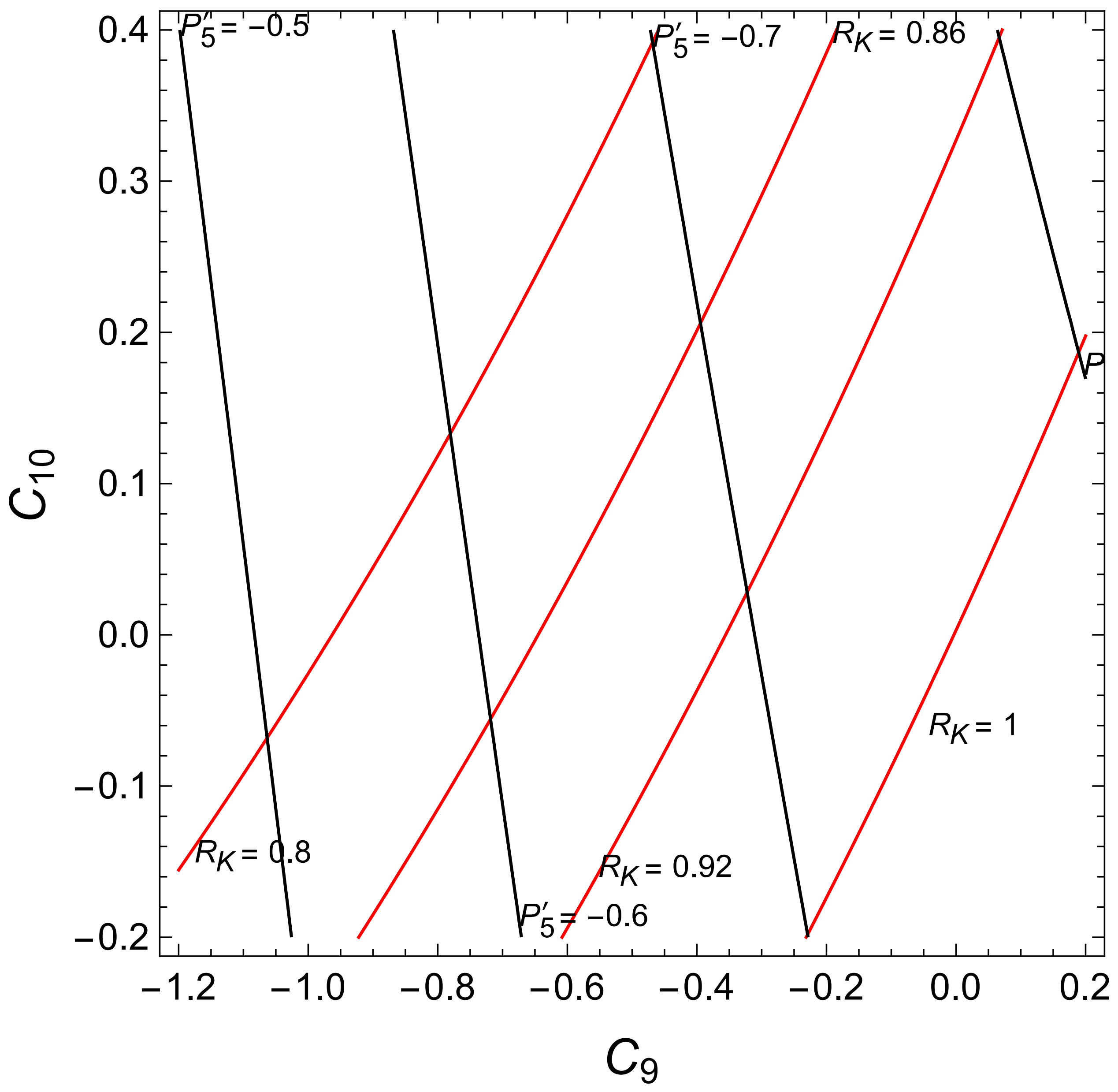




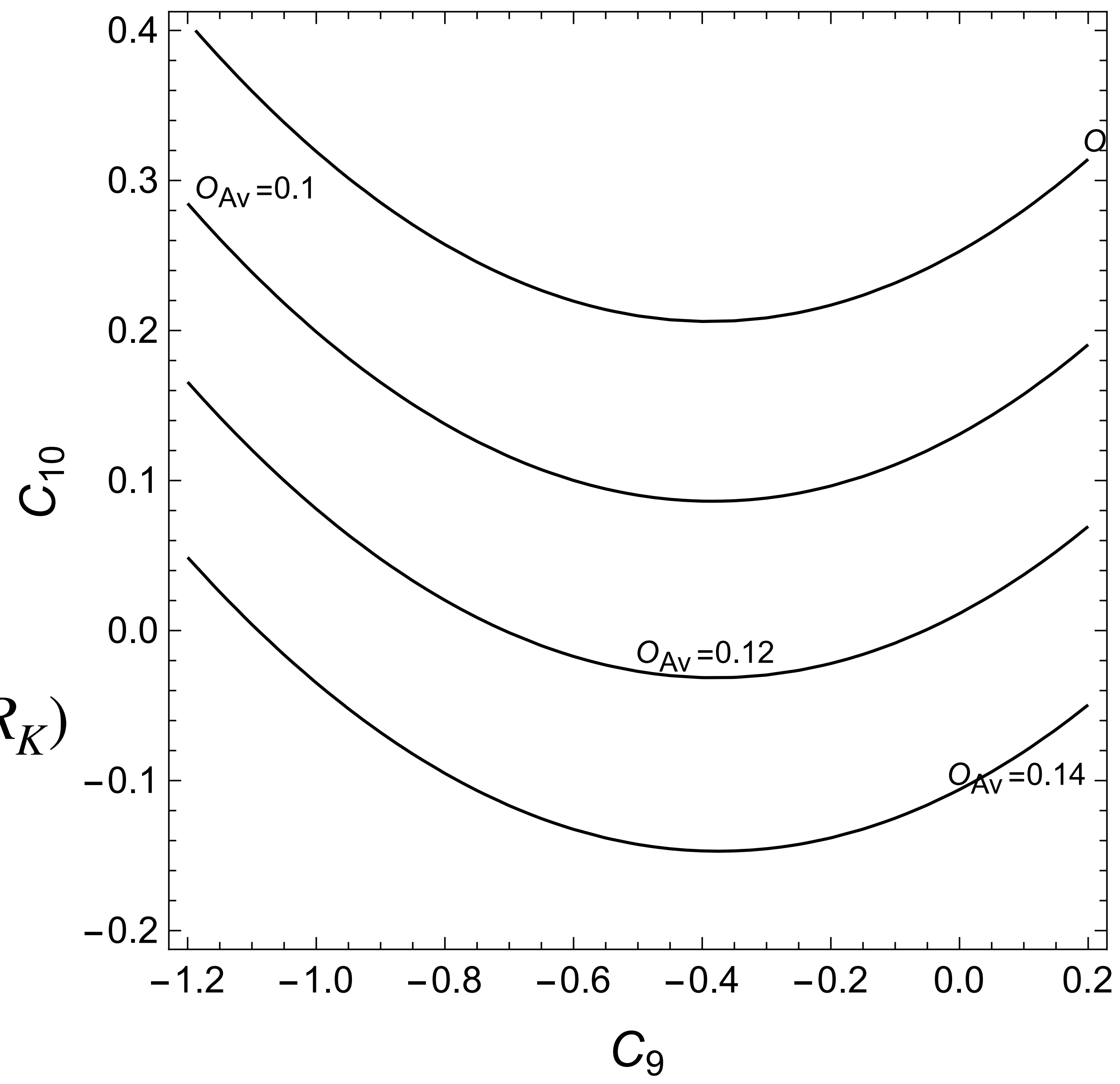
resolving power

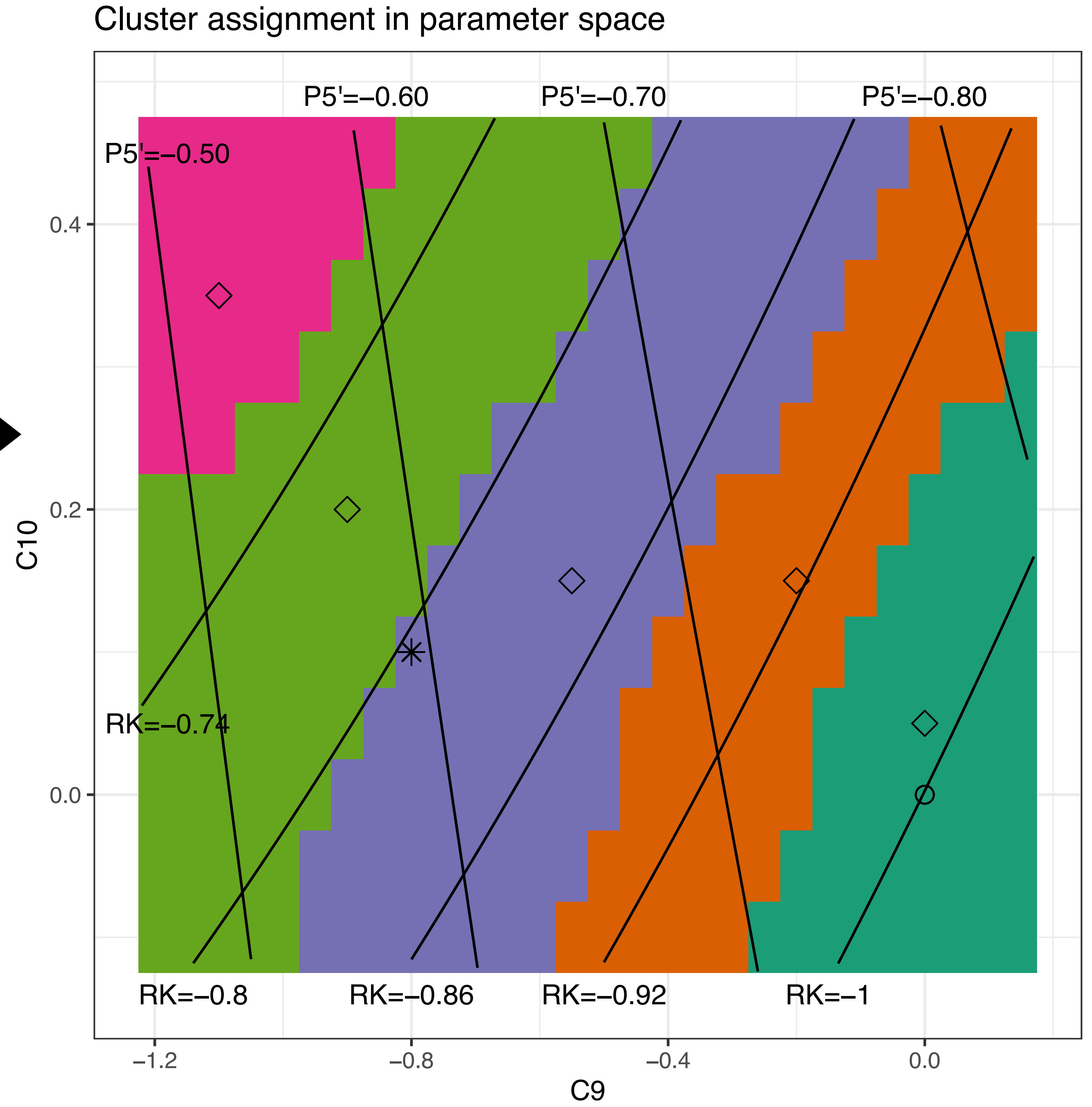
- number of partitions of parameter space with a set of measurements
- resolving power of these 14 observables with current accuracy is five clusters
- depends on
  - parameter space volume
  - range of predictions for a given observable over that region of parameter space
  - experimental and theoretical uncertainty (and correlations)
- can increase with more observables and/or better precision
  - sample of Belle II measurements with 50 ab^{-1} increases it to six
  - adding ~100 observables also used in the global fits increases it to nine

functional dependence of observables in parameter space



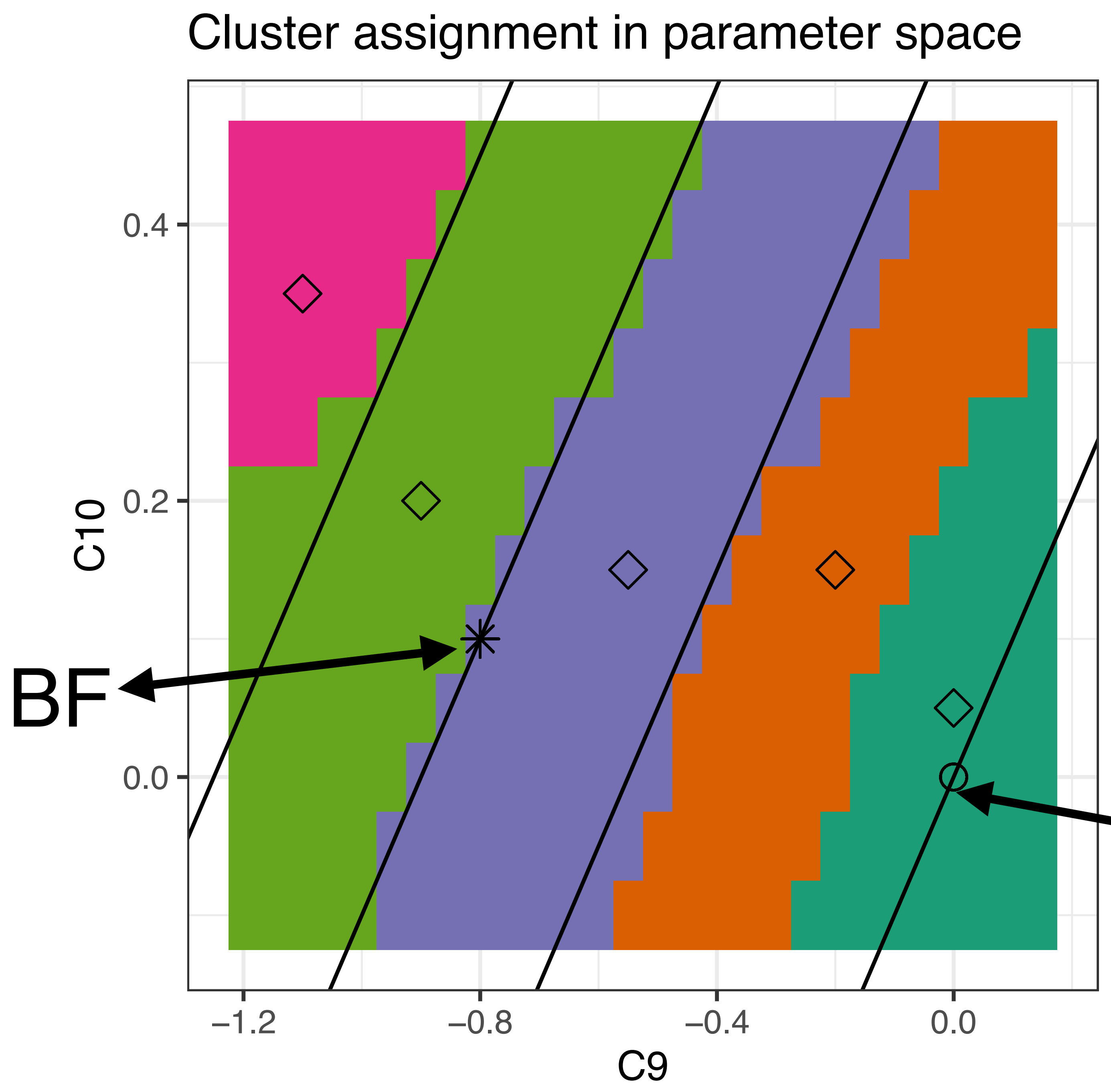
$$O_{AV} = \frac{1}{2}(P'_5 + R_K)$$

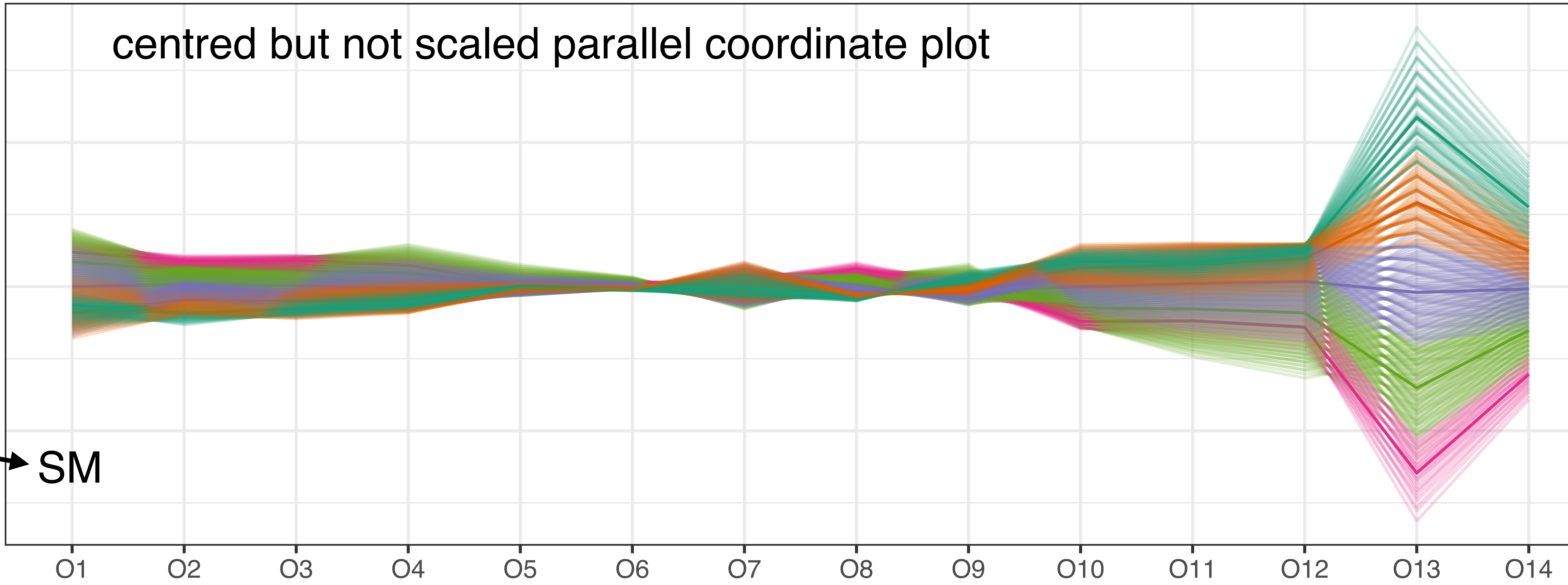




- each observable has a different functional dependence on parameters
- the figure shows 2 observables and a combination
- clustering combines all of them to visualise the collective pattern
- observables can be combined with different weights for different purposes

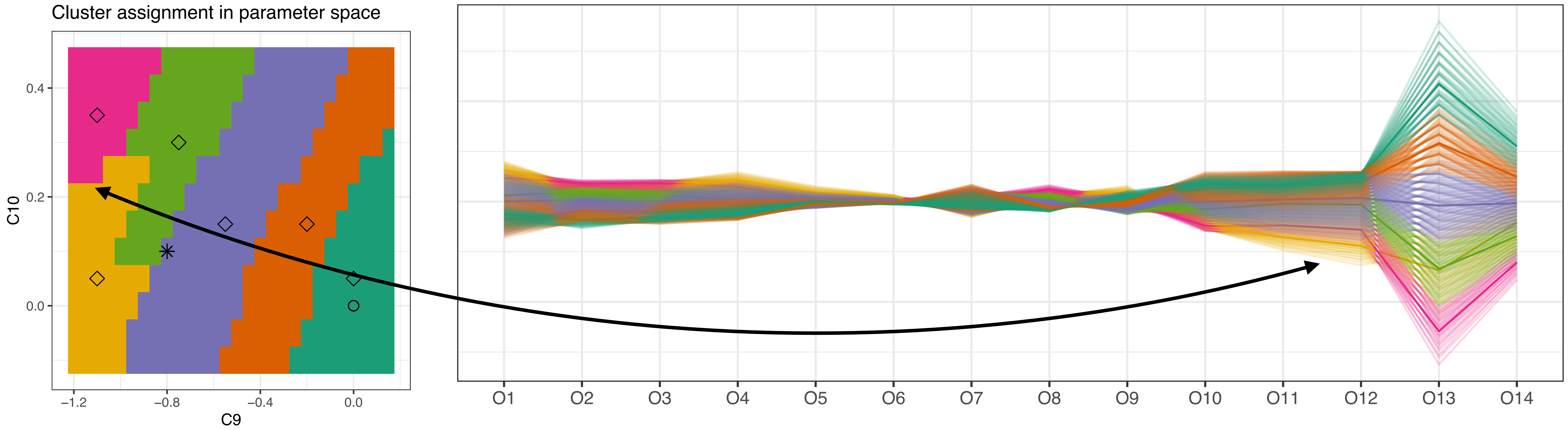
parameter vs observable space





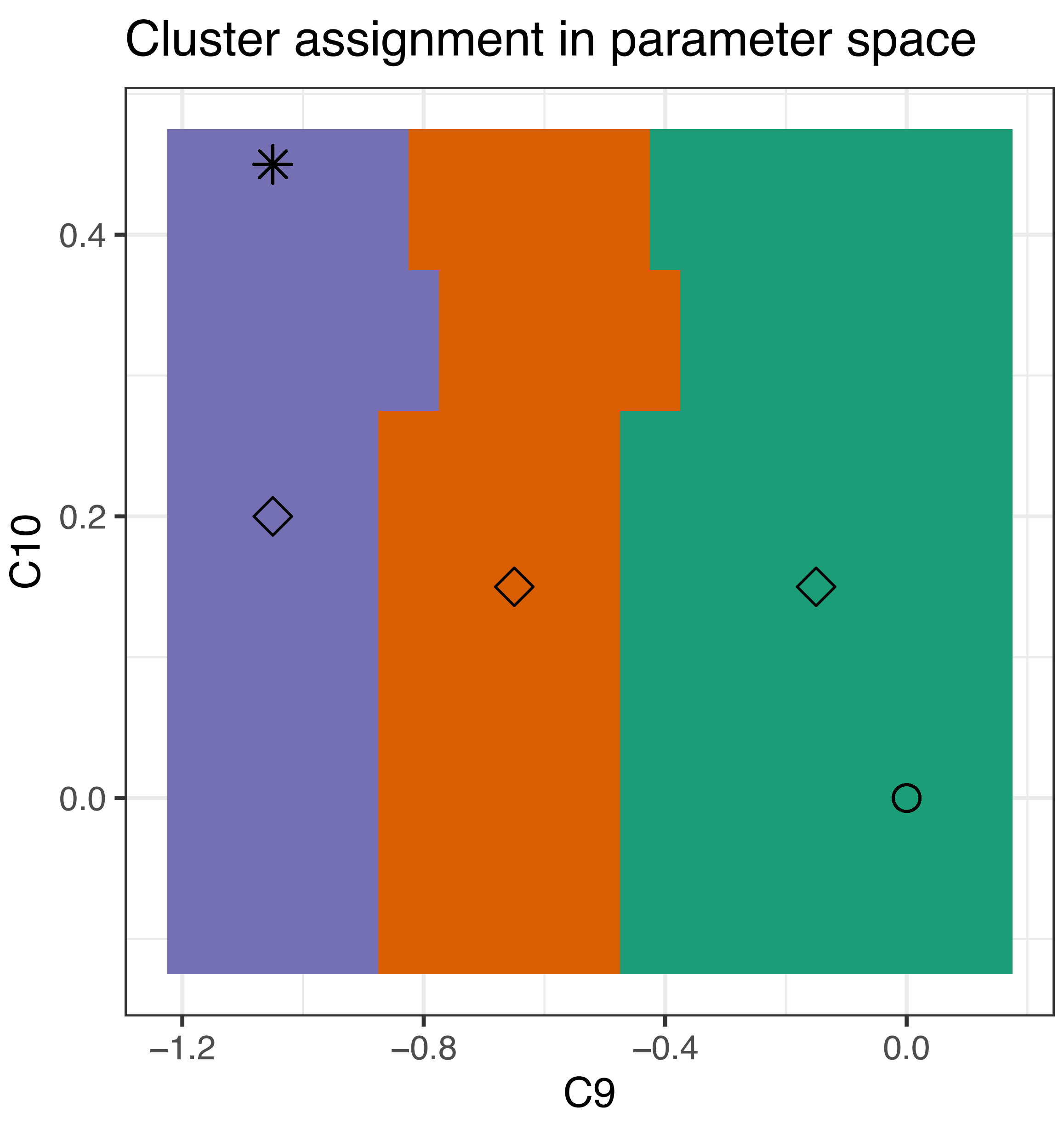
- Map the clusters in the two spaces: they are clearly separated by observables 2, 8, 13 and 14
- 13 (R_K) plays a dominant role because its predicted value (in pull units) varies the most across the parameter region (its experimental central value is not important for this)
- If an operator is dominant, as R_K is in this case, it completely determines the inter cluster boundary shape.

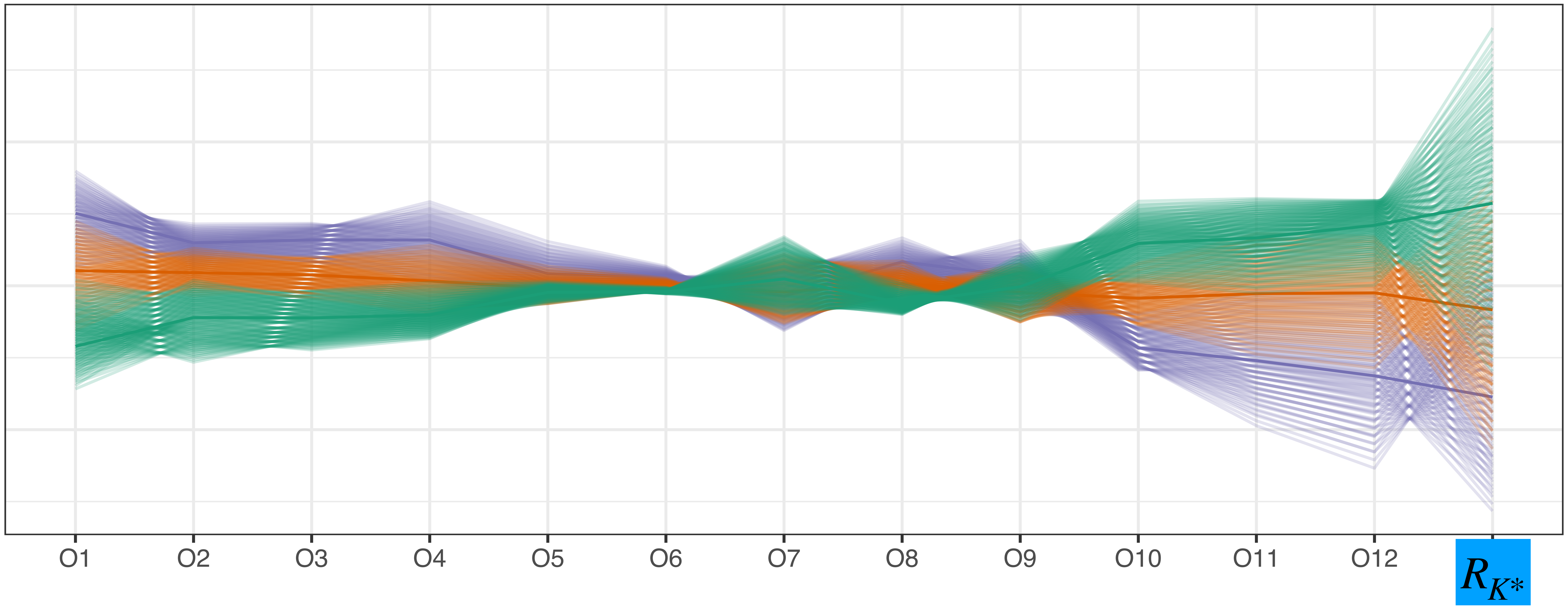
sub-leading effects



- if we add a sixth cluster: a horizontal partition (C_{10} sensitivity) appears far from the SM
- $O_{11,12}$ or $P_2[6-8]$, $P_2[15-19]$ are important for yellow-pink boundary
- caveat: numerical accuracy affects small details

remove R_K

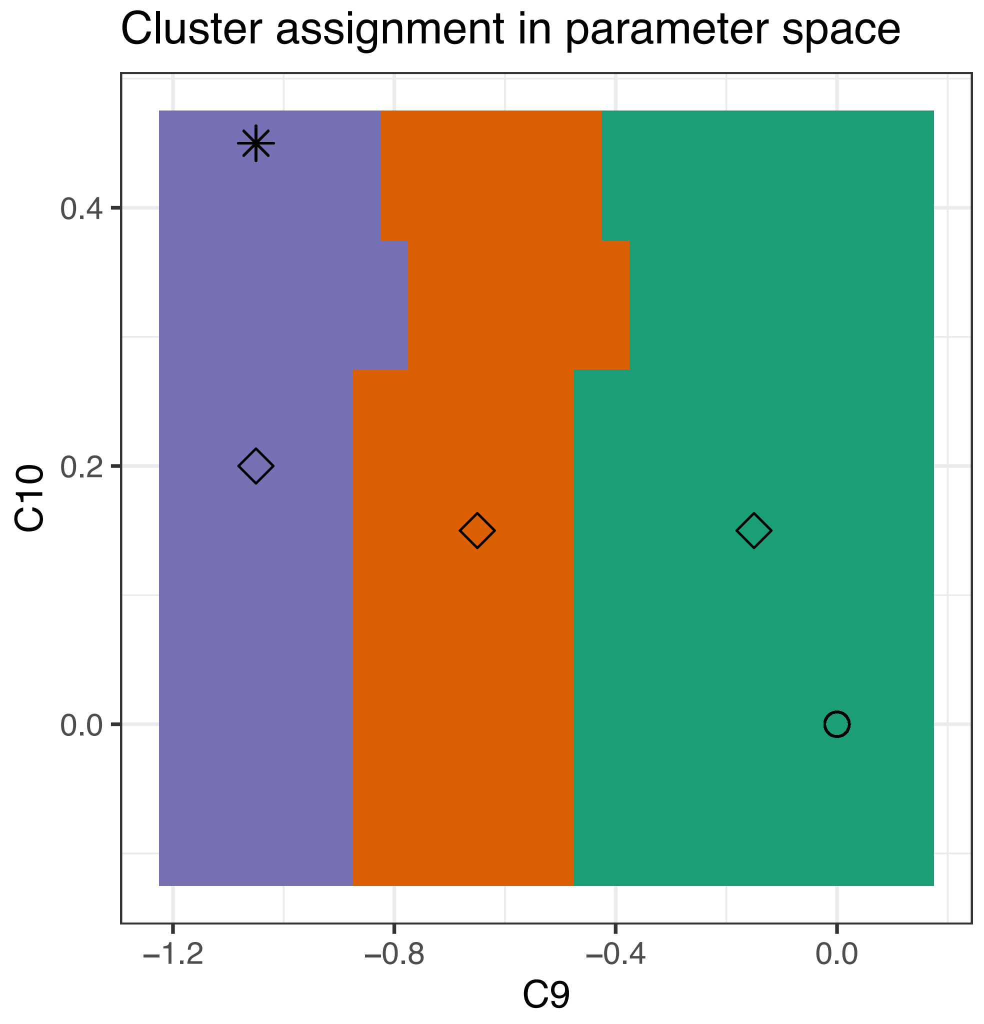


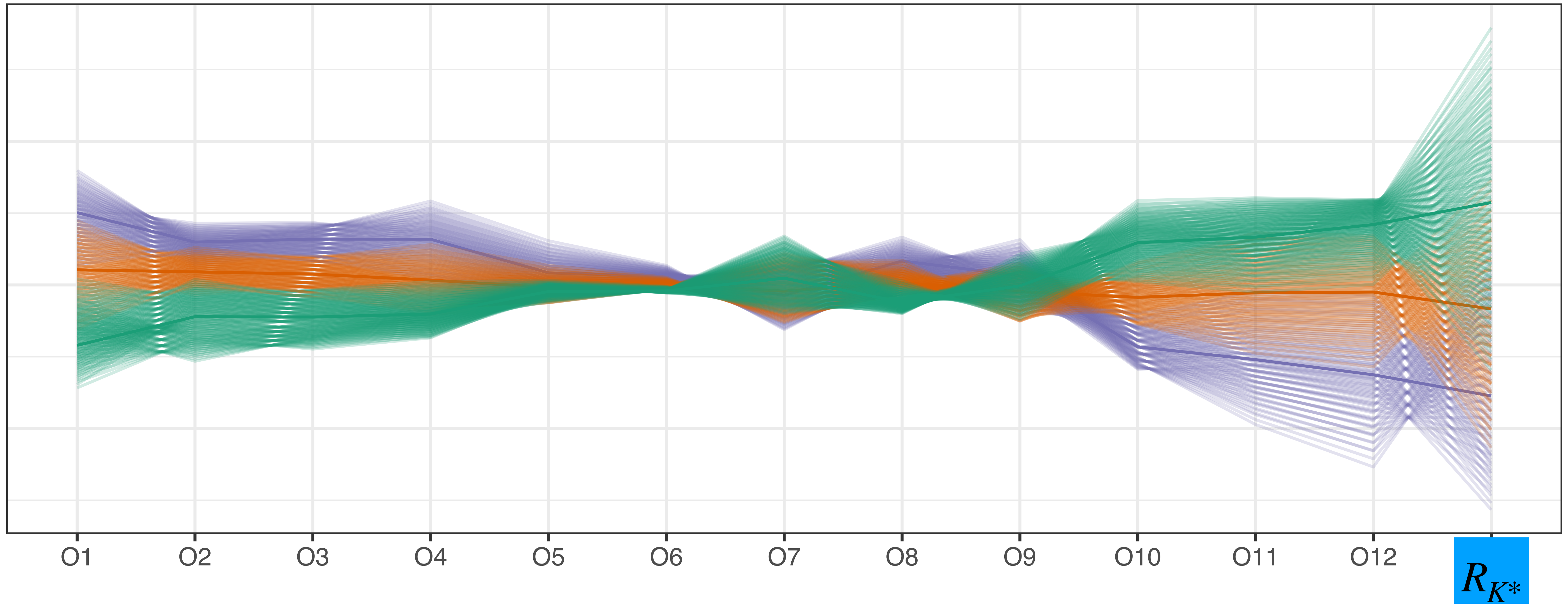


- resolving power reduced to about 3 clusters, partitioning mostly along C_9
- no one observable is responsible for cluster boundaries, it is a collective effect
- separation of brown cluster more related to P_5' (see cluster overlap in P_2)
- notice change in BF position as well

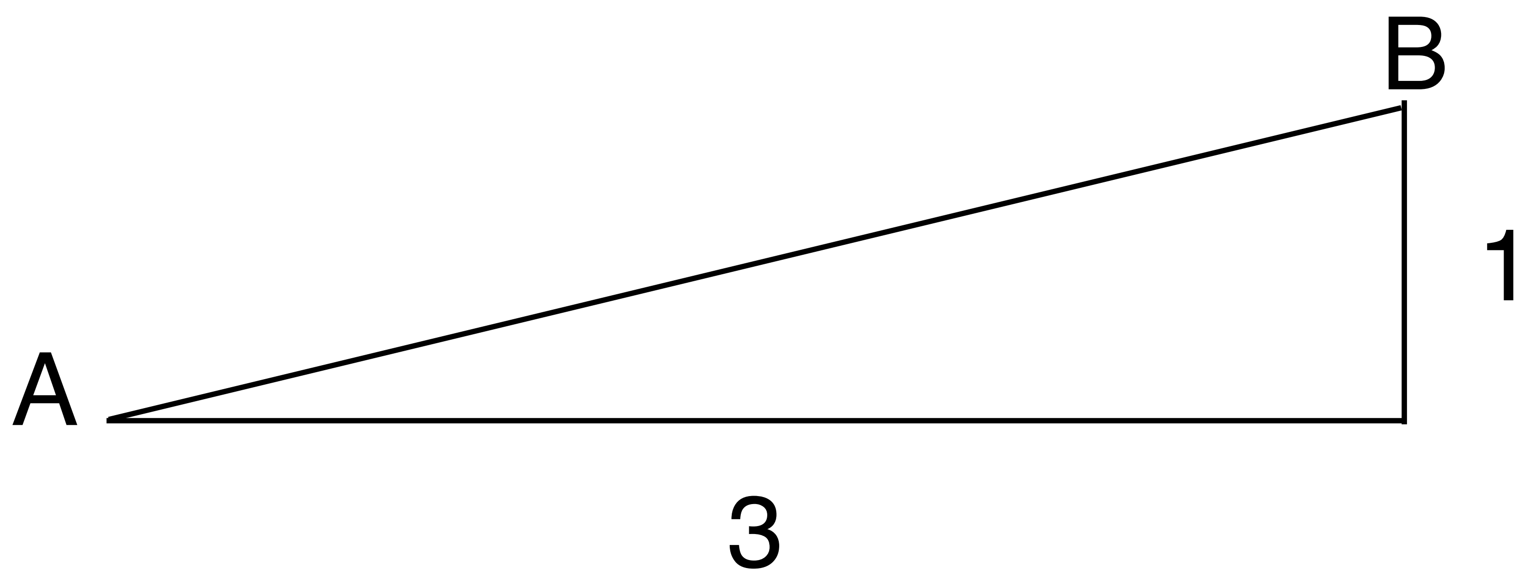
increasing the importance of dominant observables

- three clusters without R_K Euclidean distance



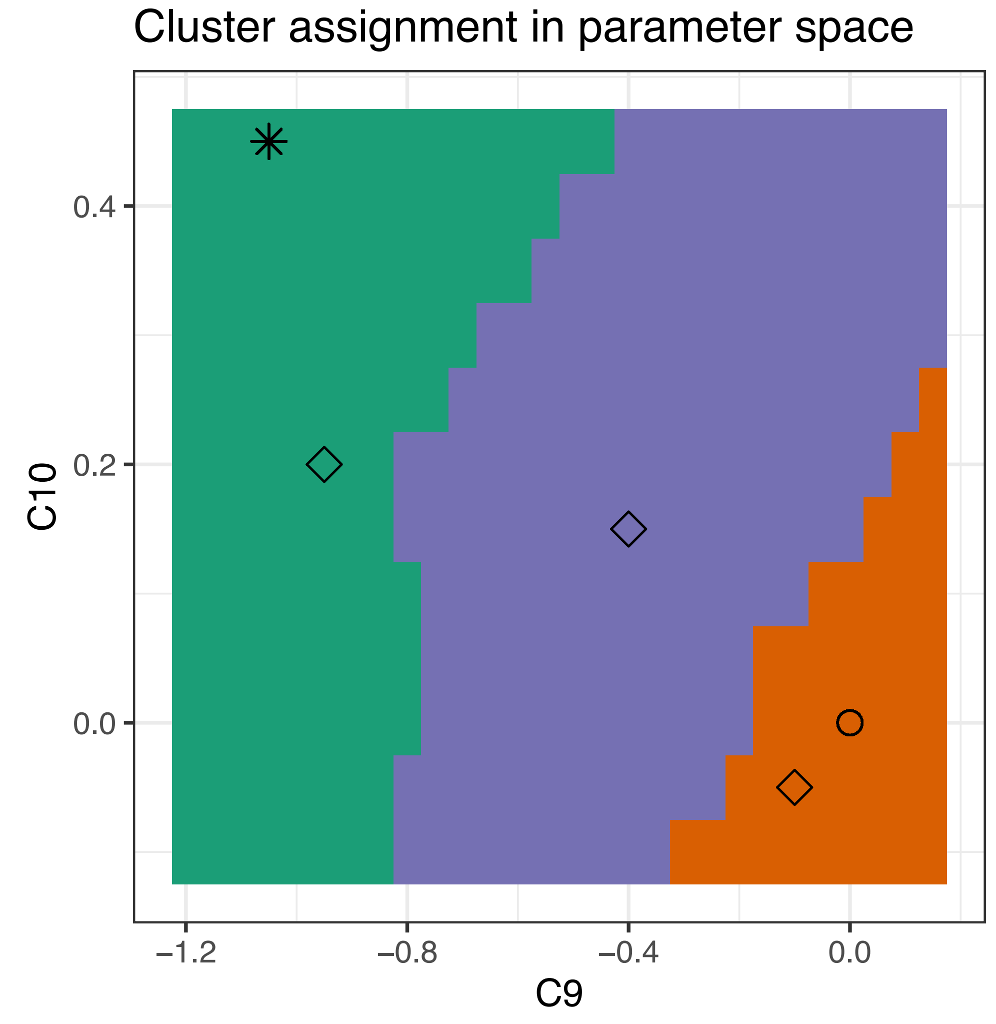


- Using maximum distance with complete linkage we increase the importance of R_K^*:

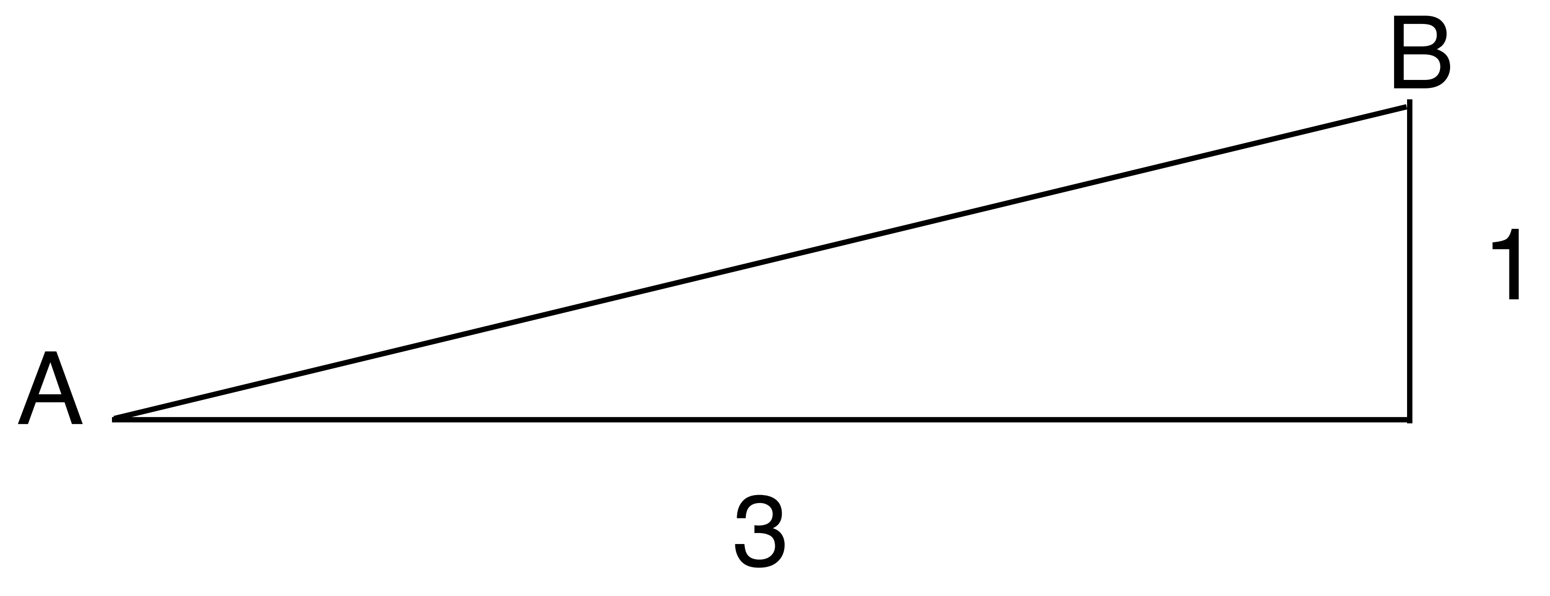


$$d_{\text{Chebyshev}}(A, B) = 3$$

$$d_{\text{Euclidean}}(A, B) = 3.16$$



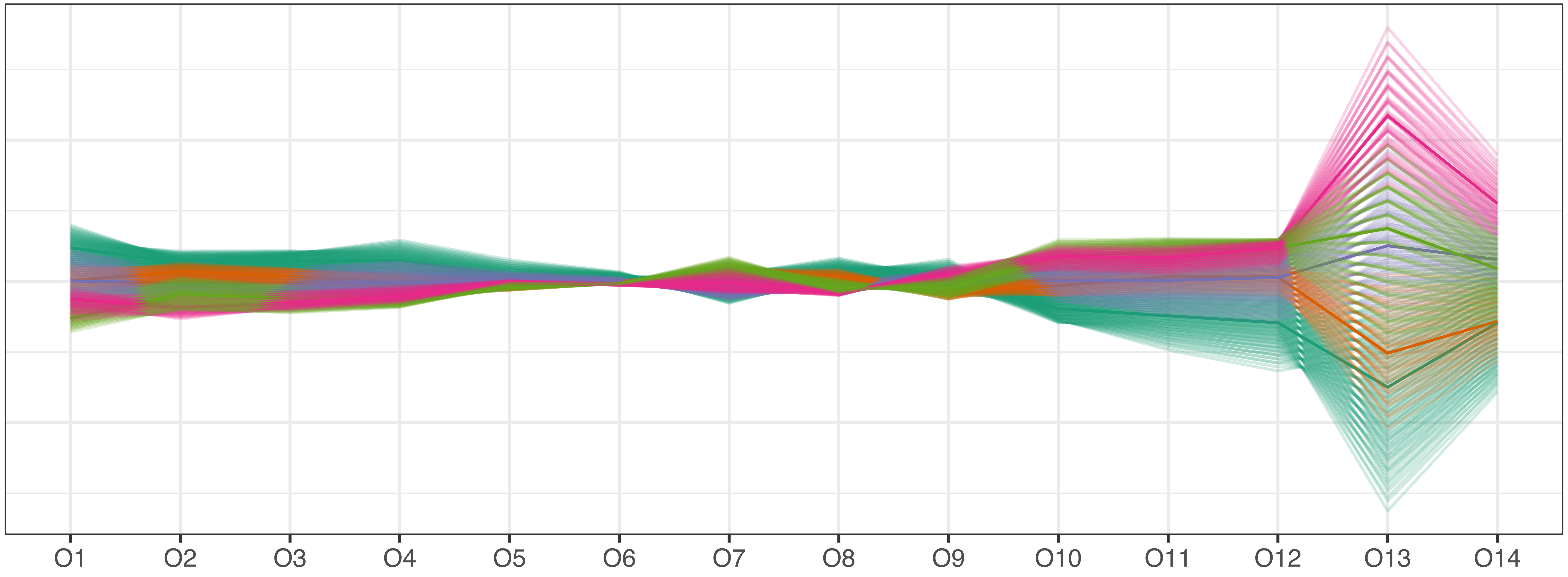
increasing the importance of sub-dominant observables without removing any

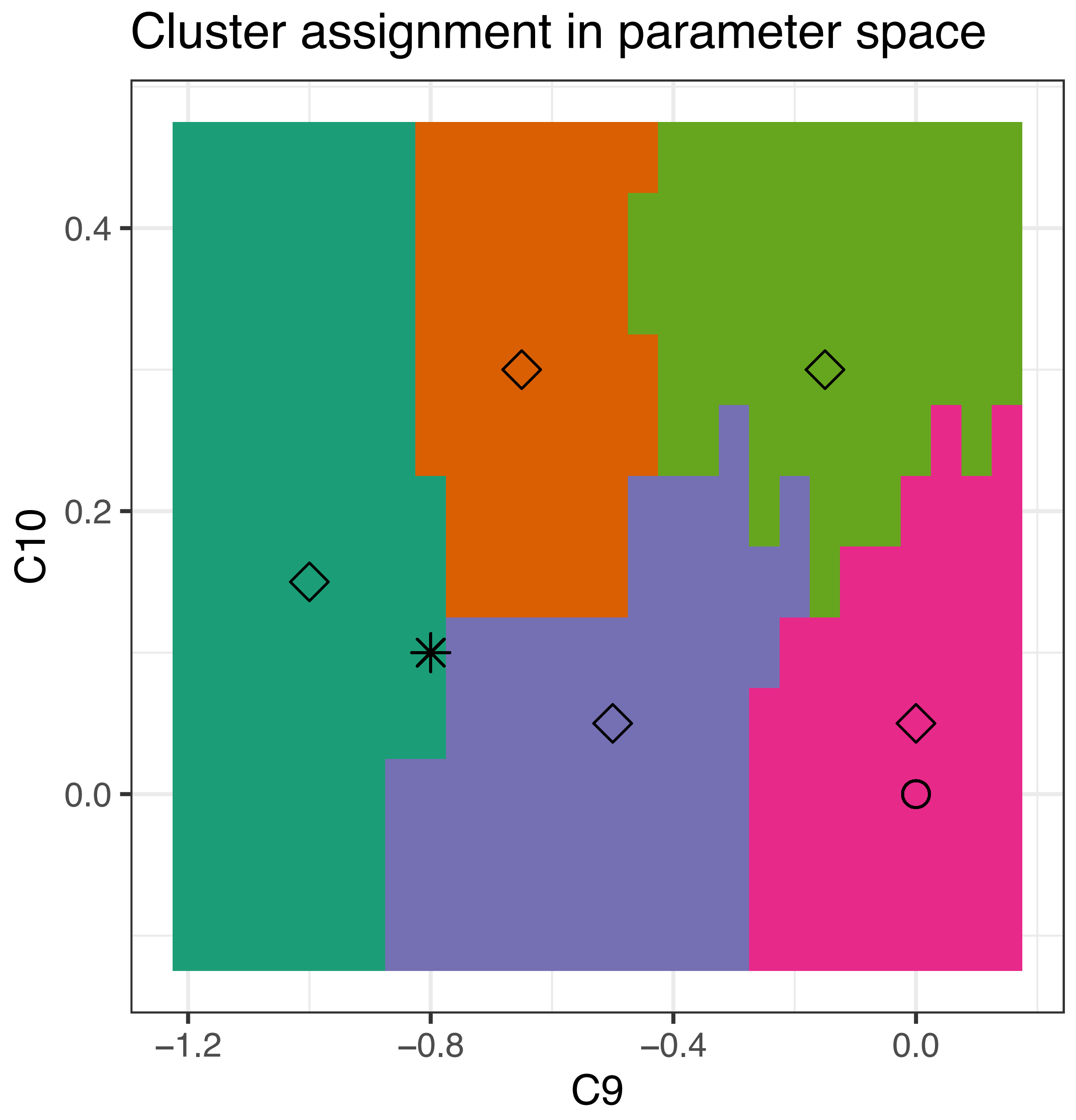


$$d_{\text{Manhattan}}(A, B) = 4$$

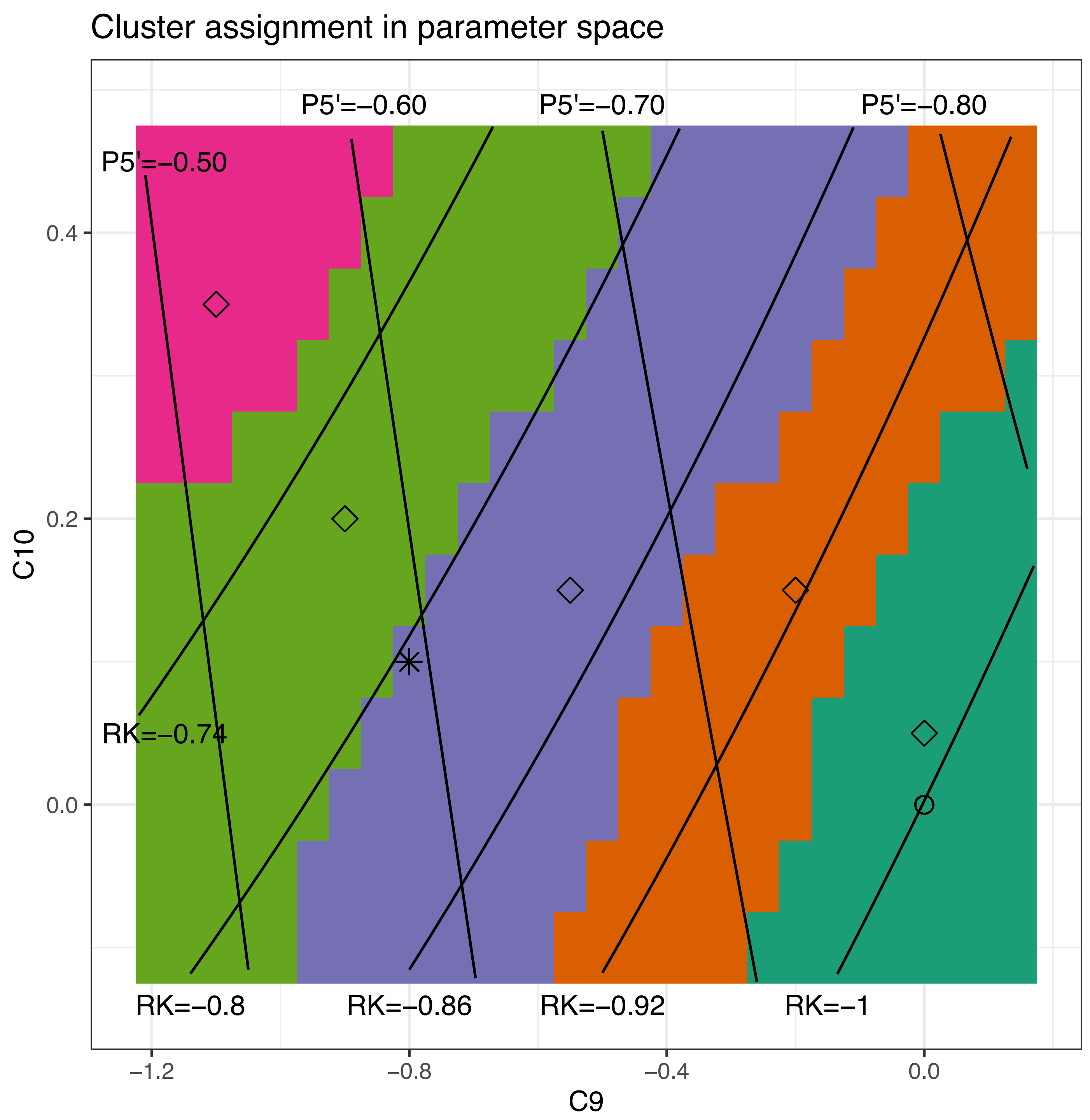
$$d_{\text{Euclidean}}(A, B) = 3.16$$

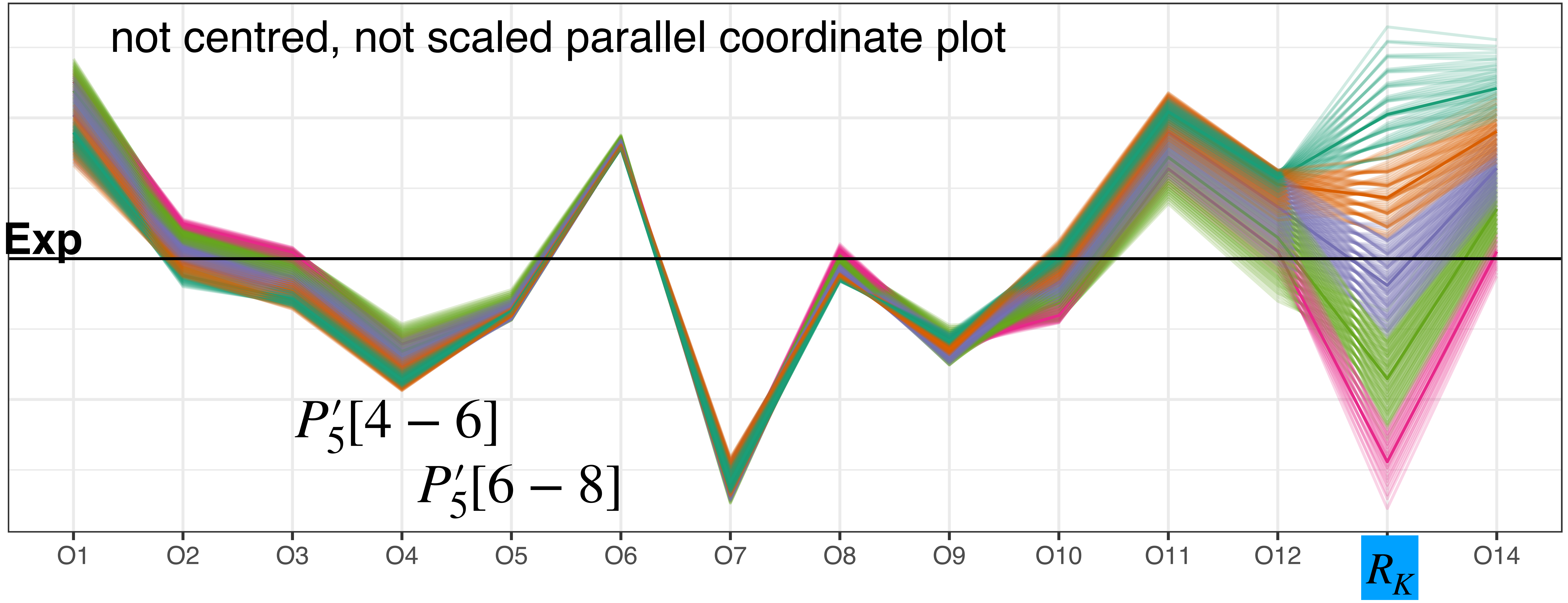
- Manhattan distance, Ward linkage
- boundary shapes collective effect
- see PC for R_K





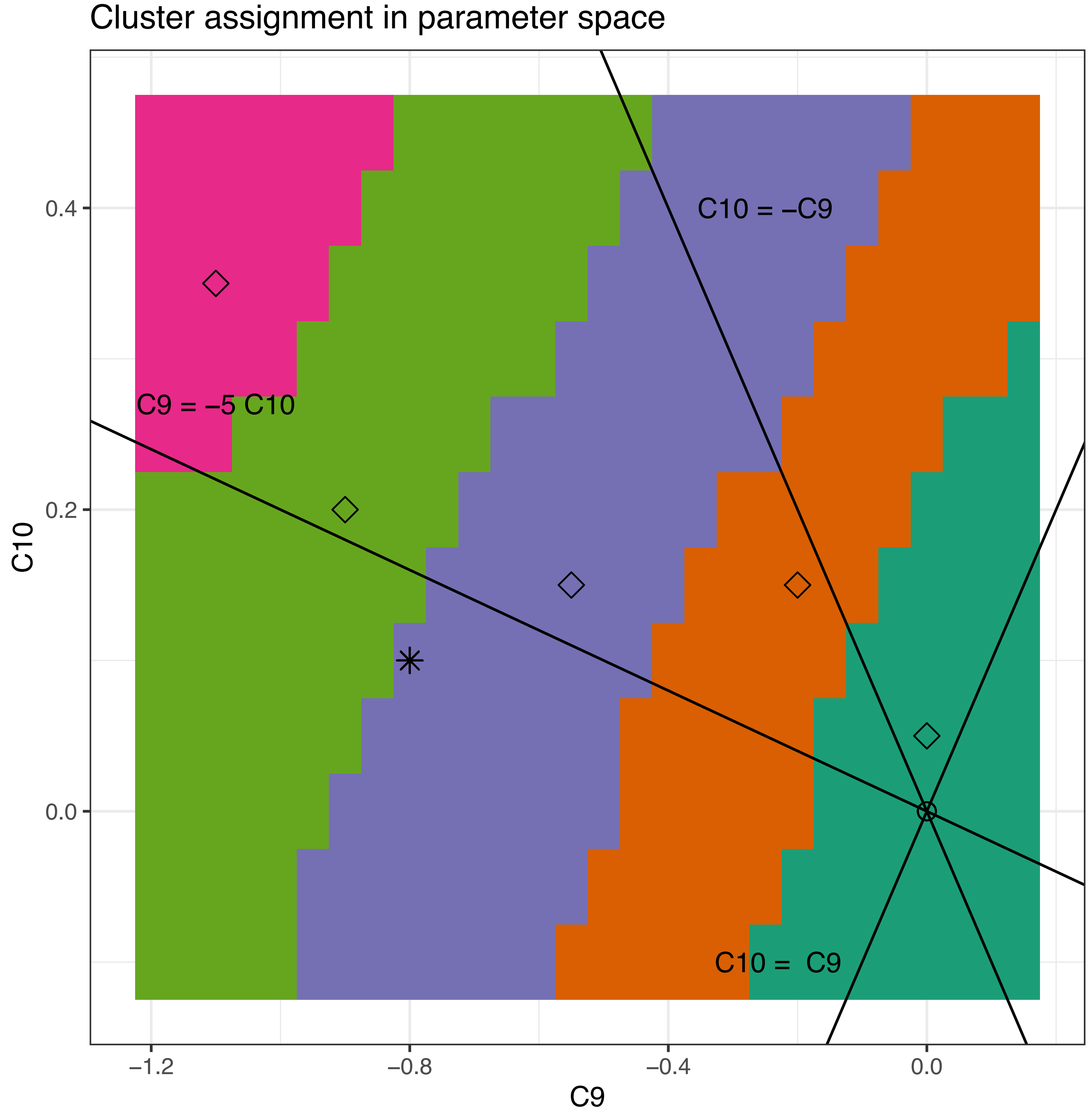
internal tensions in the fit: P_5' vs R_K



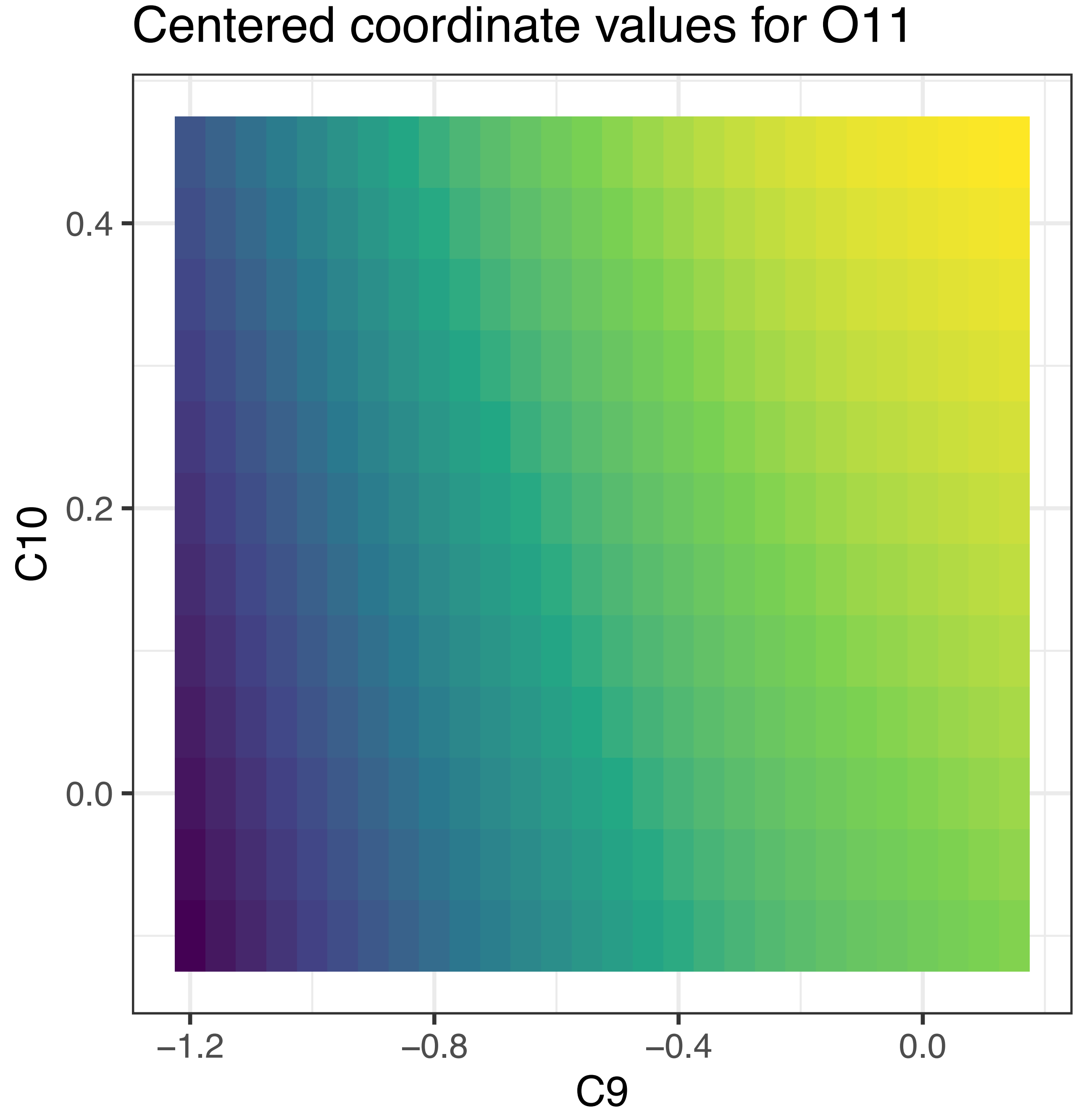


- BF on the boundary between light green and purple clusters
- $P_5'[4-6]$, $P_5'[6-8]$ (largest discrepancy between SM and BF in P_5') prefer the light green: larger negative C_9 (recall the experimental value $P_5'[4-6] = -0.39 \pm 0.11$)
- $R_K = 0.86 \pm 0.06$ prefers the purple cluster (but not R_{K*})
- the model points that take $P_5'[4-6]$, $P_5'[6-8]$ closest to experiment, take R_K furthest away

sensitivity along certain directions

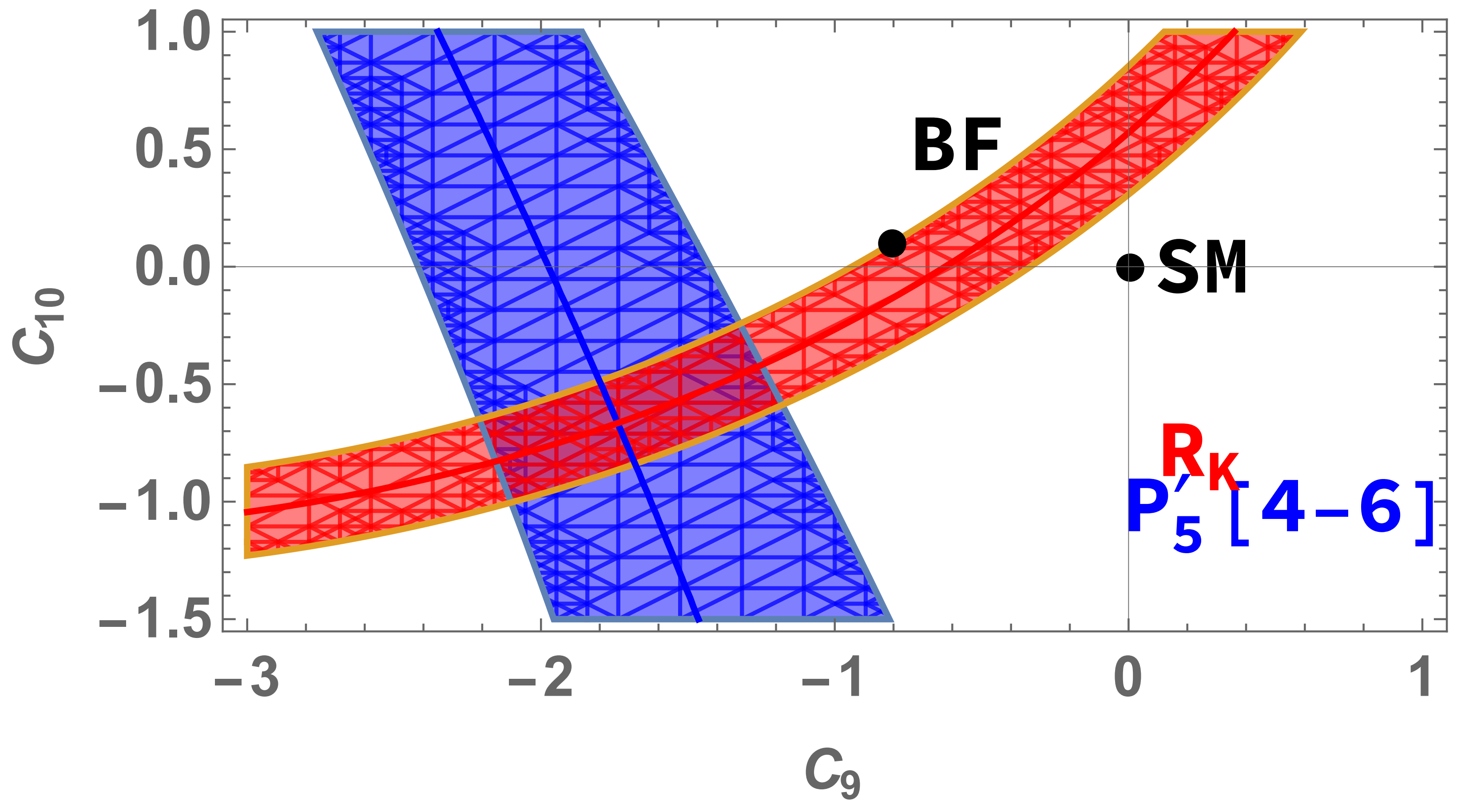


- most sensitive to a direction along $C_{10} \approx 0.2\ C_9$
- almost no sensitivity along $C_{10} = C_9$
- these patterns agree with what is seen in 2d global fits
- With this set, sensitivity along $C_{10} = C_9$ can be increased with a more precise measurement of $O_{11} = P_2[6-8]$

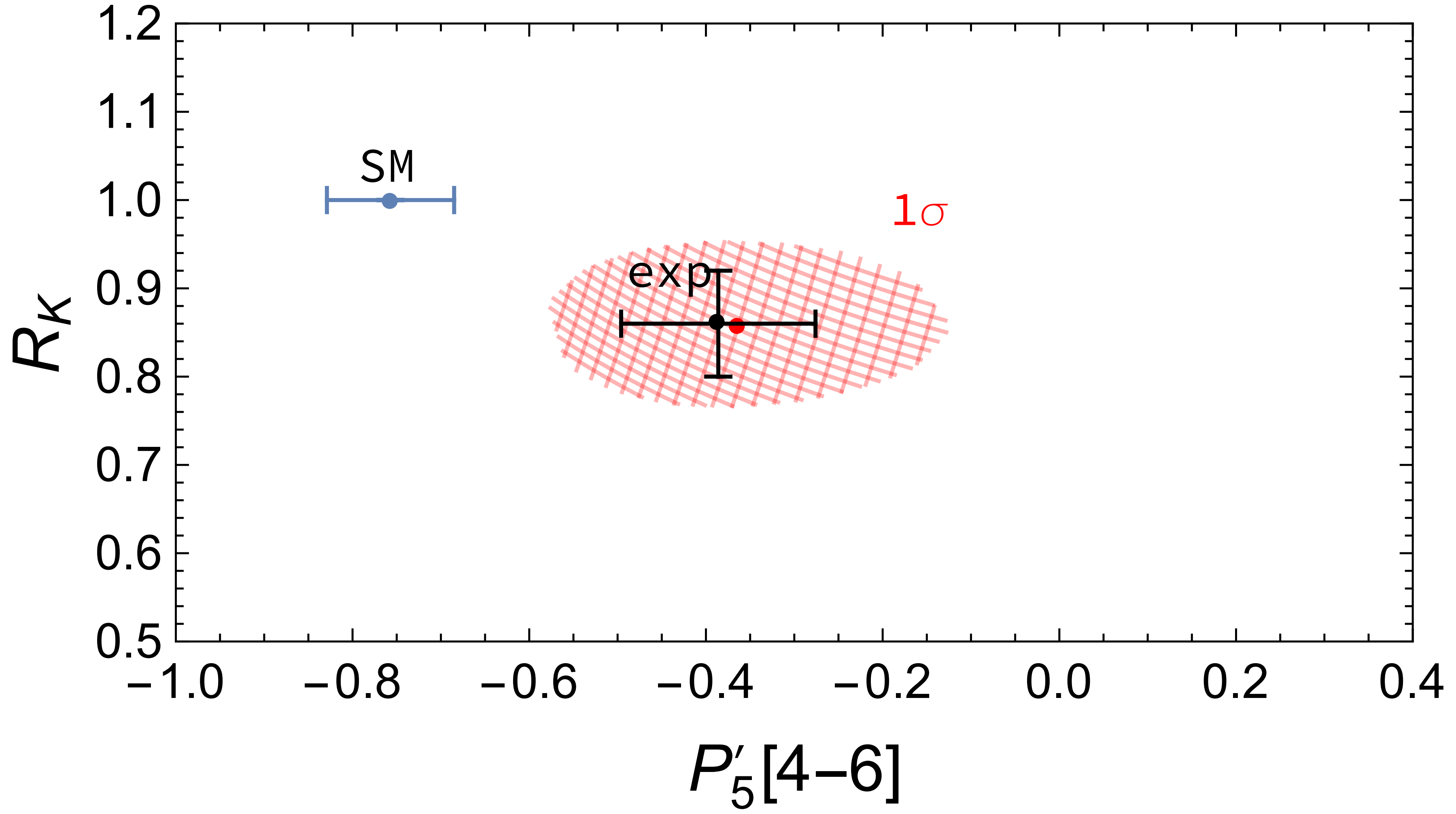


overall view in observable space

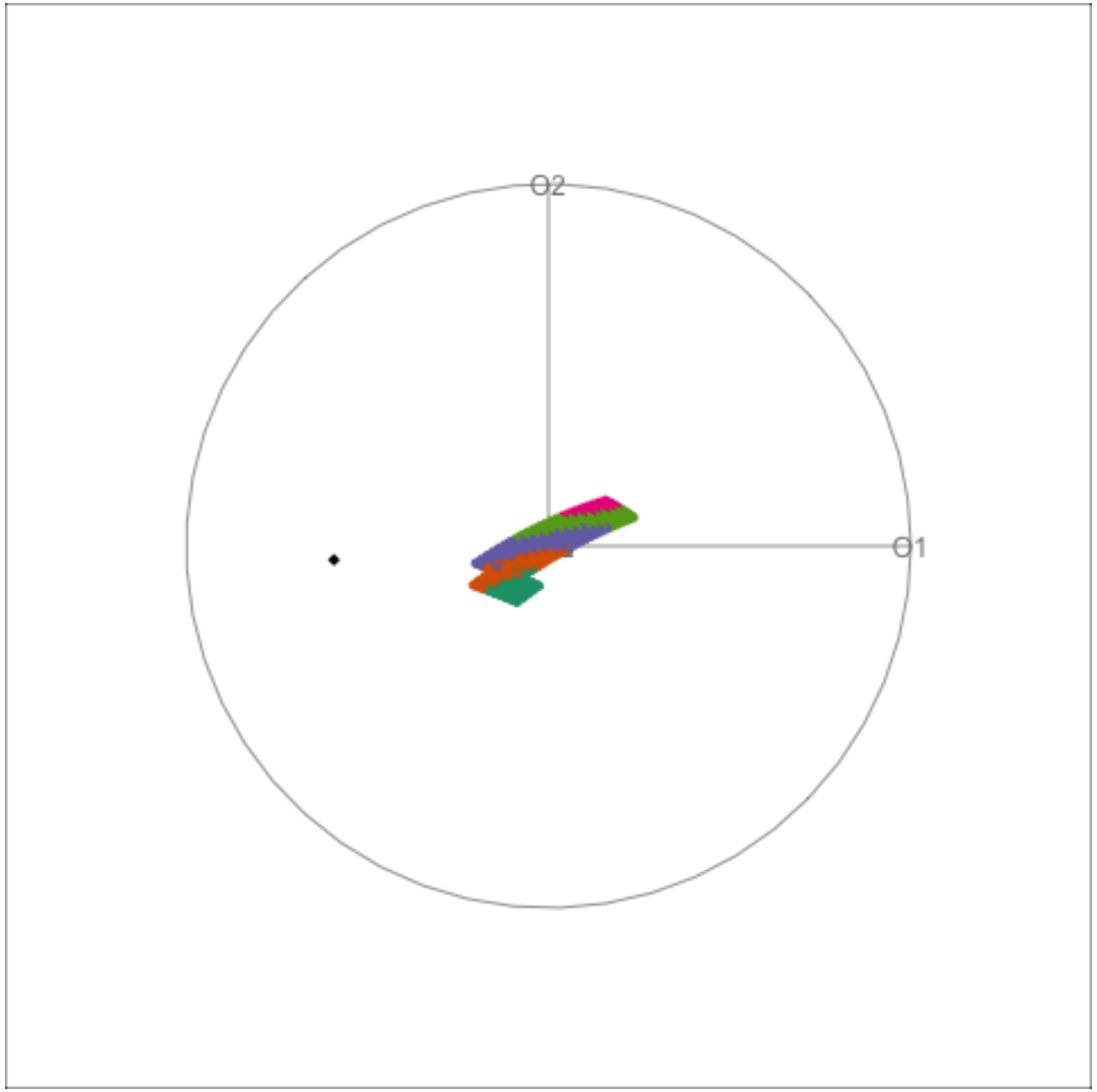
usually we see something like this



would like to see something like this

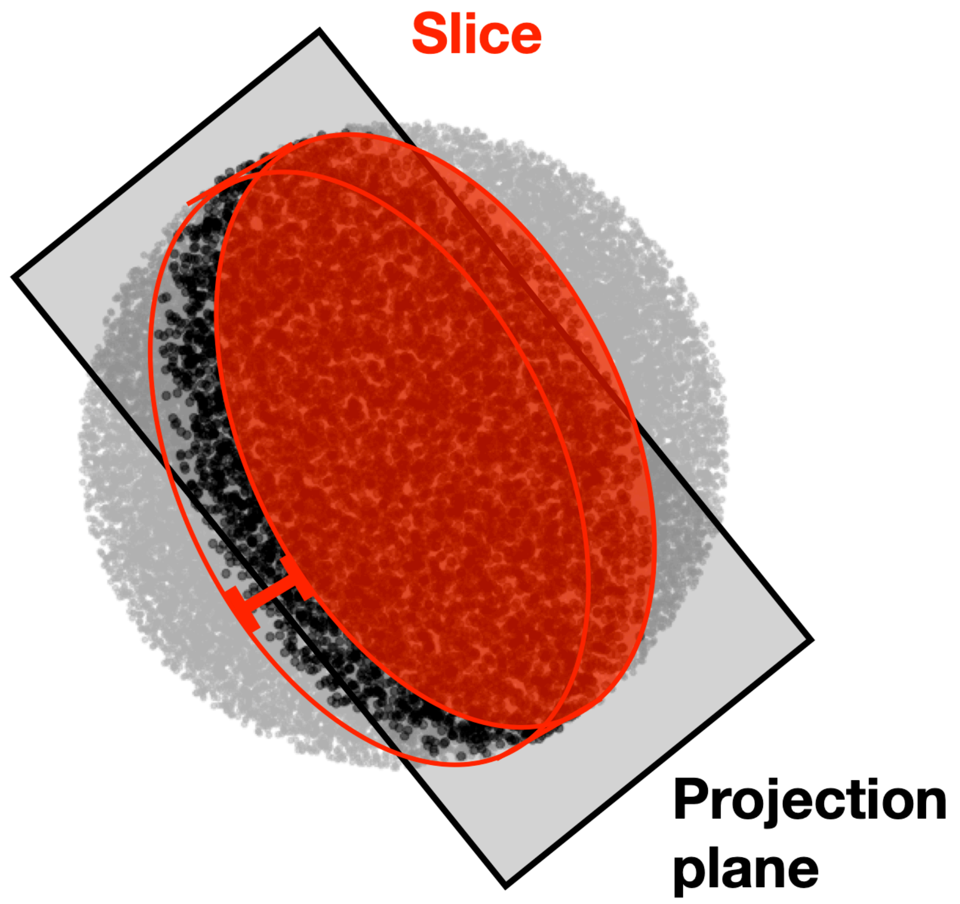


14 dimensions

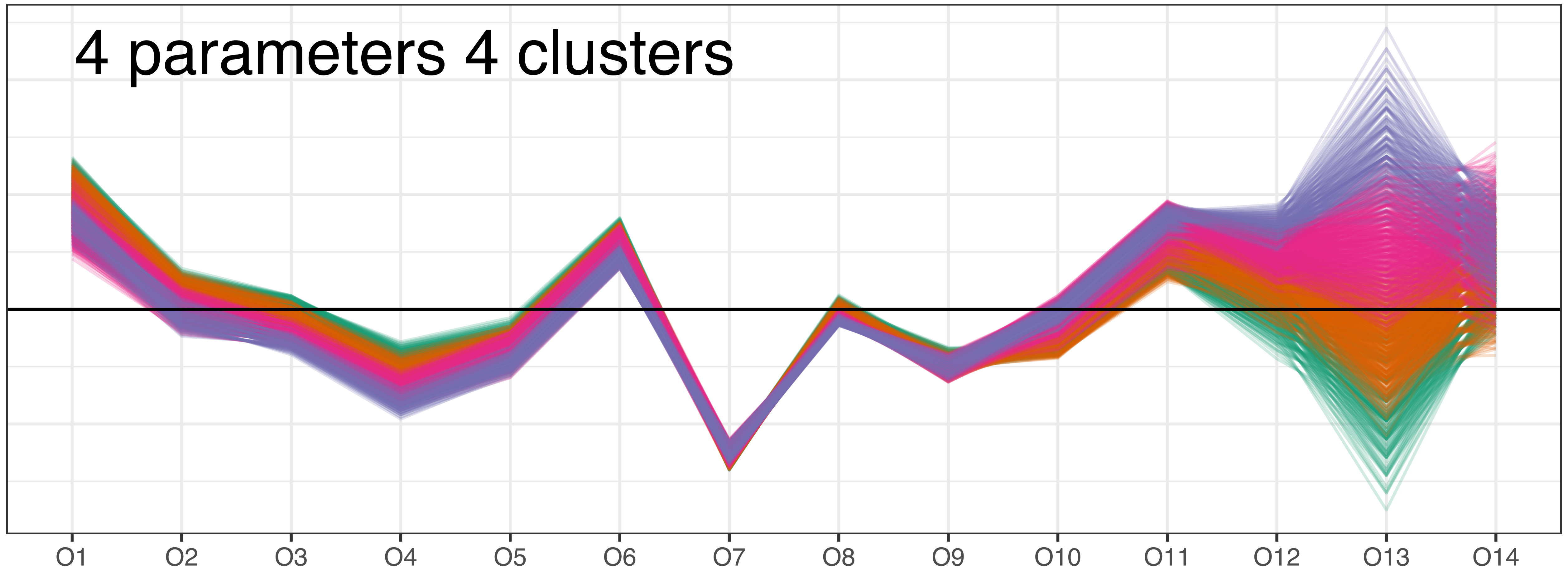


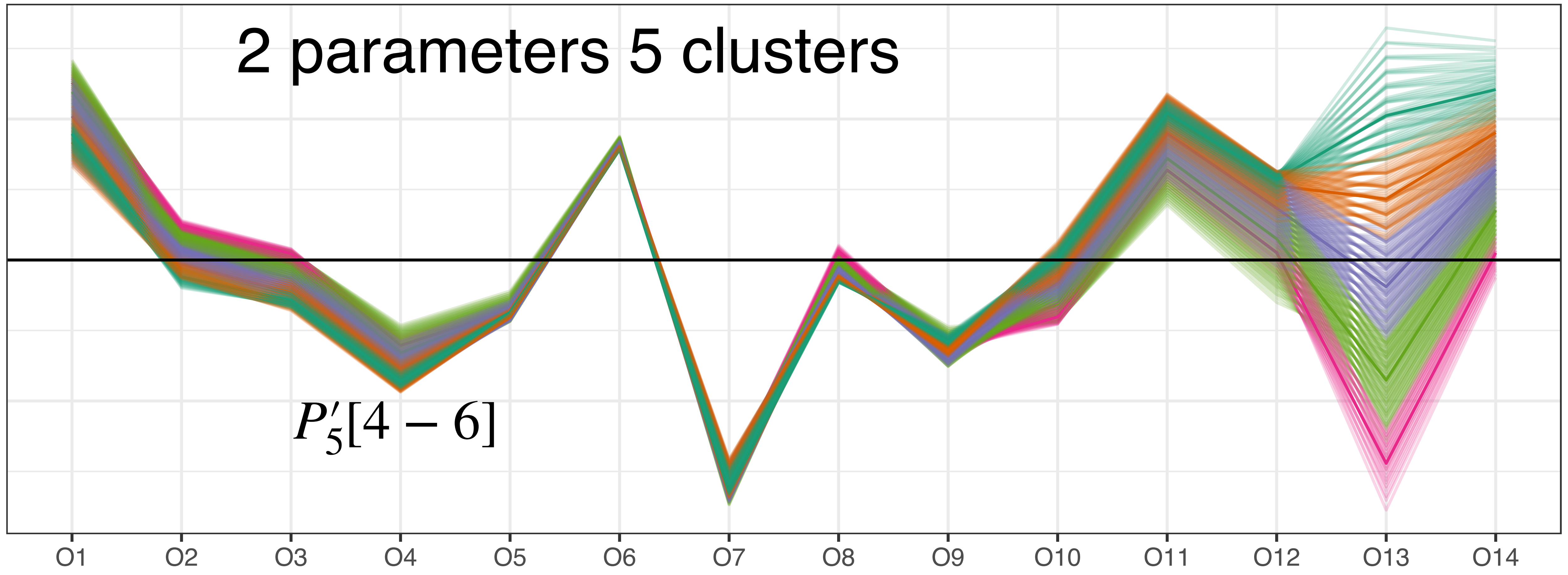
animations available at https://uschilaa.github.io/animations/ (some), or running the app interactively

increase the number of parameters to 4



include $C^{\mu}_{9'}$ and $C^{\mu}_{10'}$

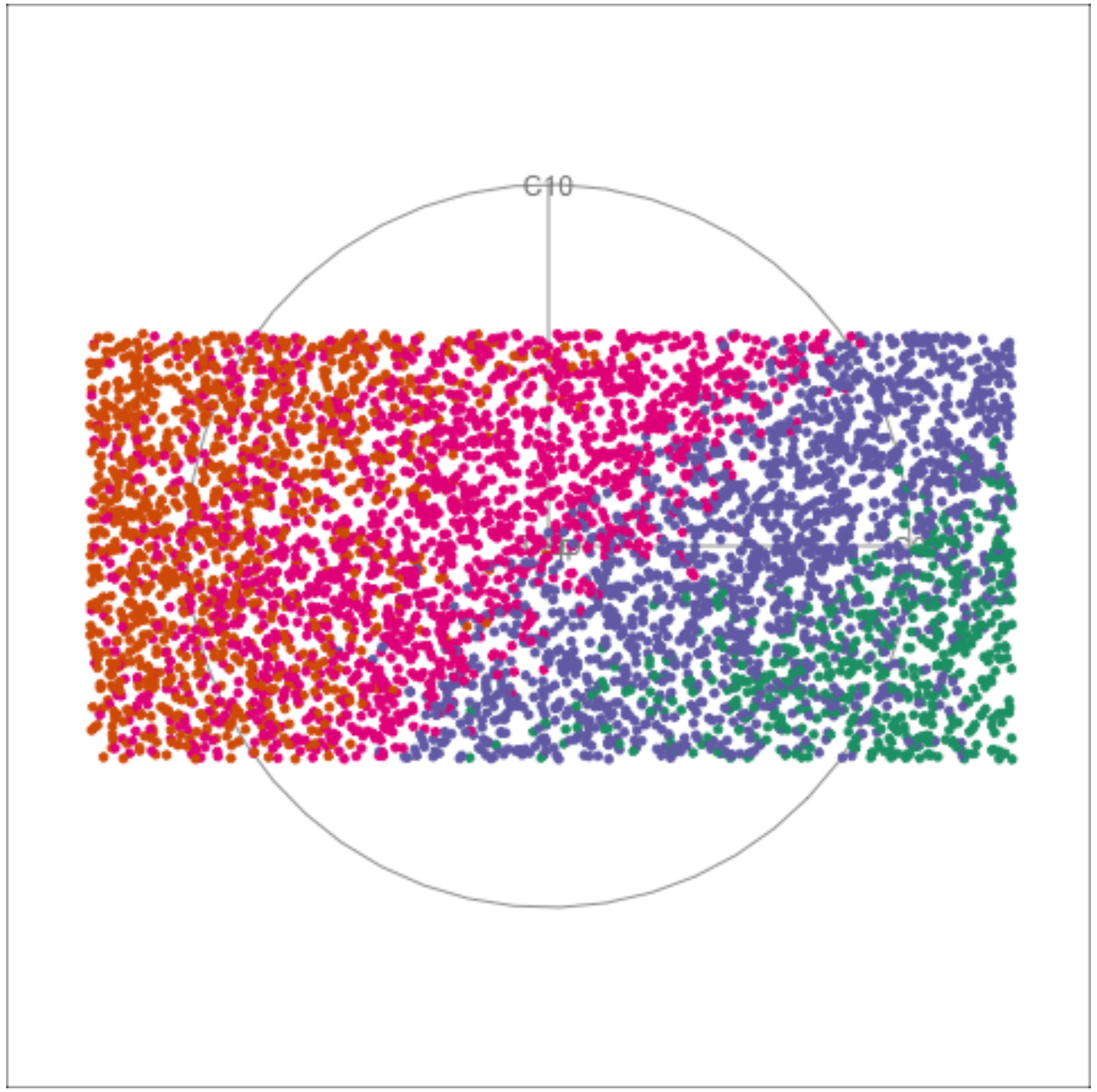




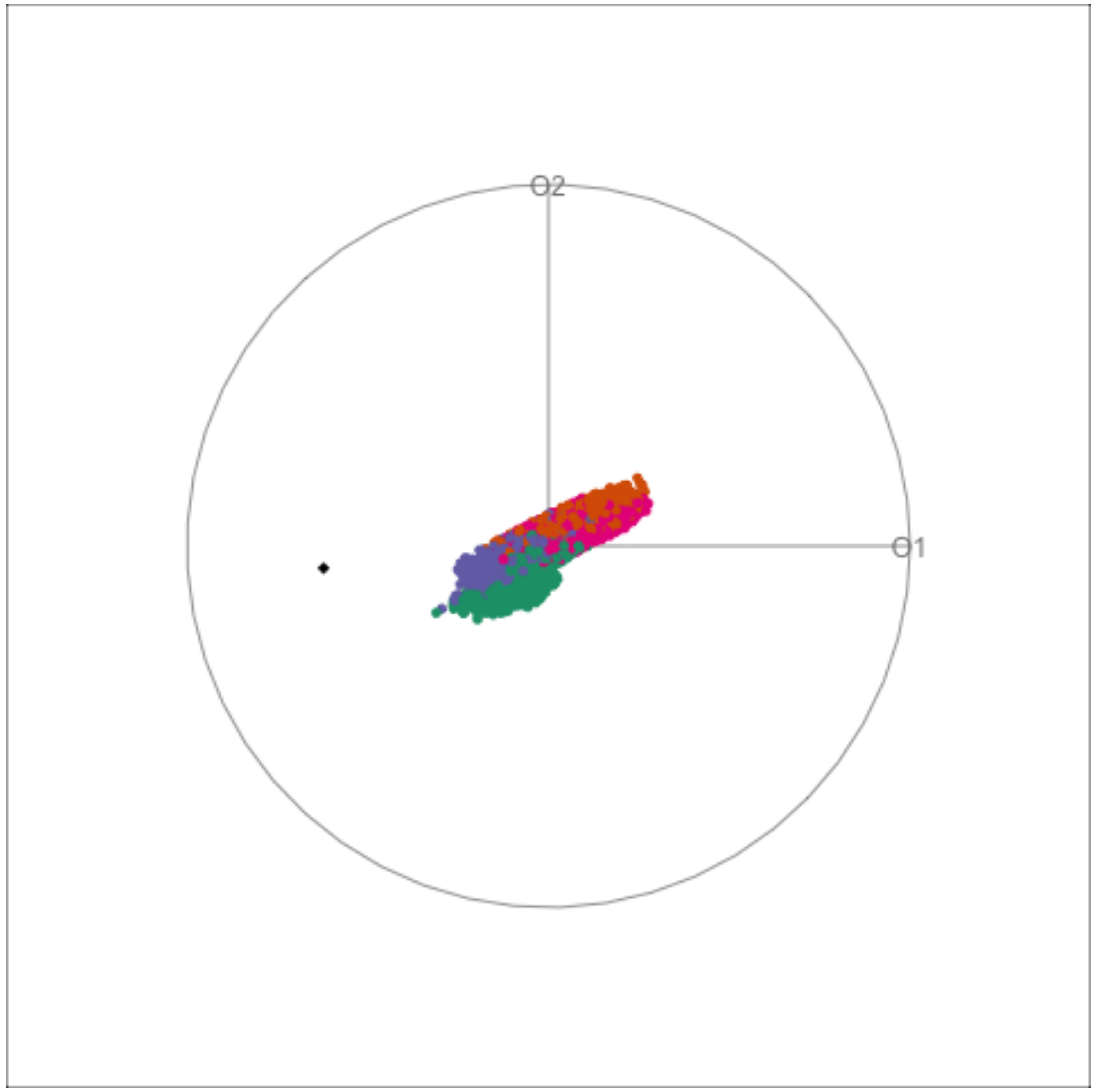
- resolution is 4 clusters
  - now 1σ cluster radius/separation for four degrees of freedom (depend on 4d volume)
- extended range of predictions with the two additional parameters increases overlap with experiment
  - compare $O_1,\ O_4,\ O_6$ for example with the 2 parameter case
- Notice how $R_K(O_{13})$ no longer cleanly separates the clusters
- Reduced tension between P'_5 and R_K

Parameter-observable space visualisation

- to view the partitioning of parameter space with more than two parameters requires visualising a second high-dimensional space (with slicing)

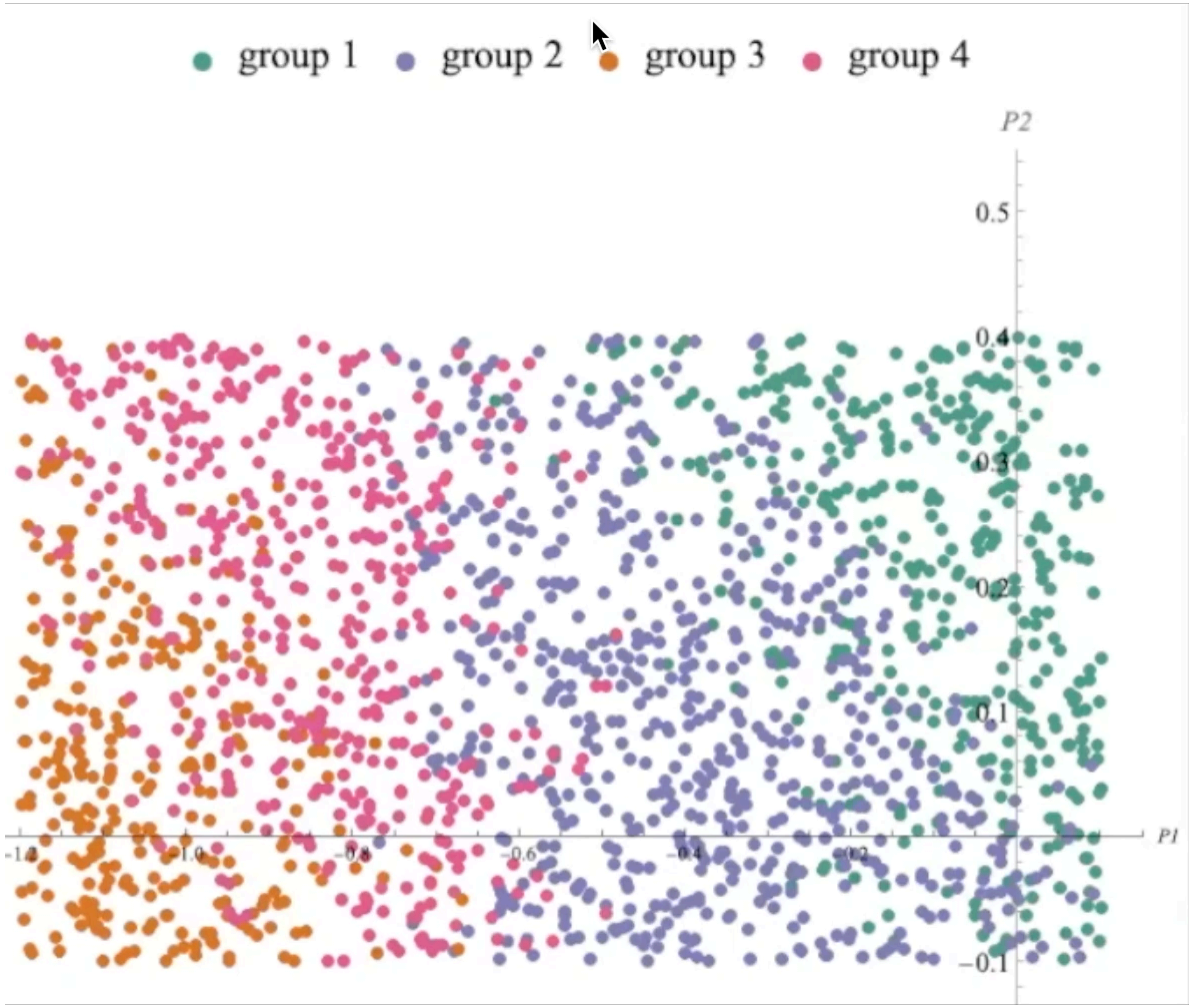


tour in parameter space

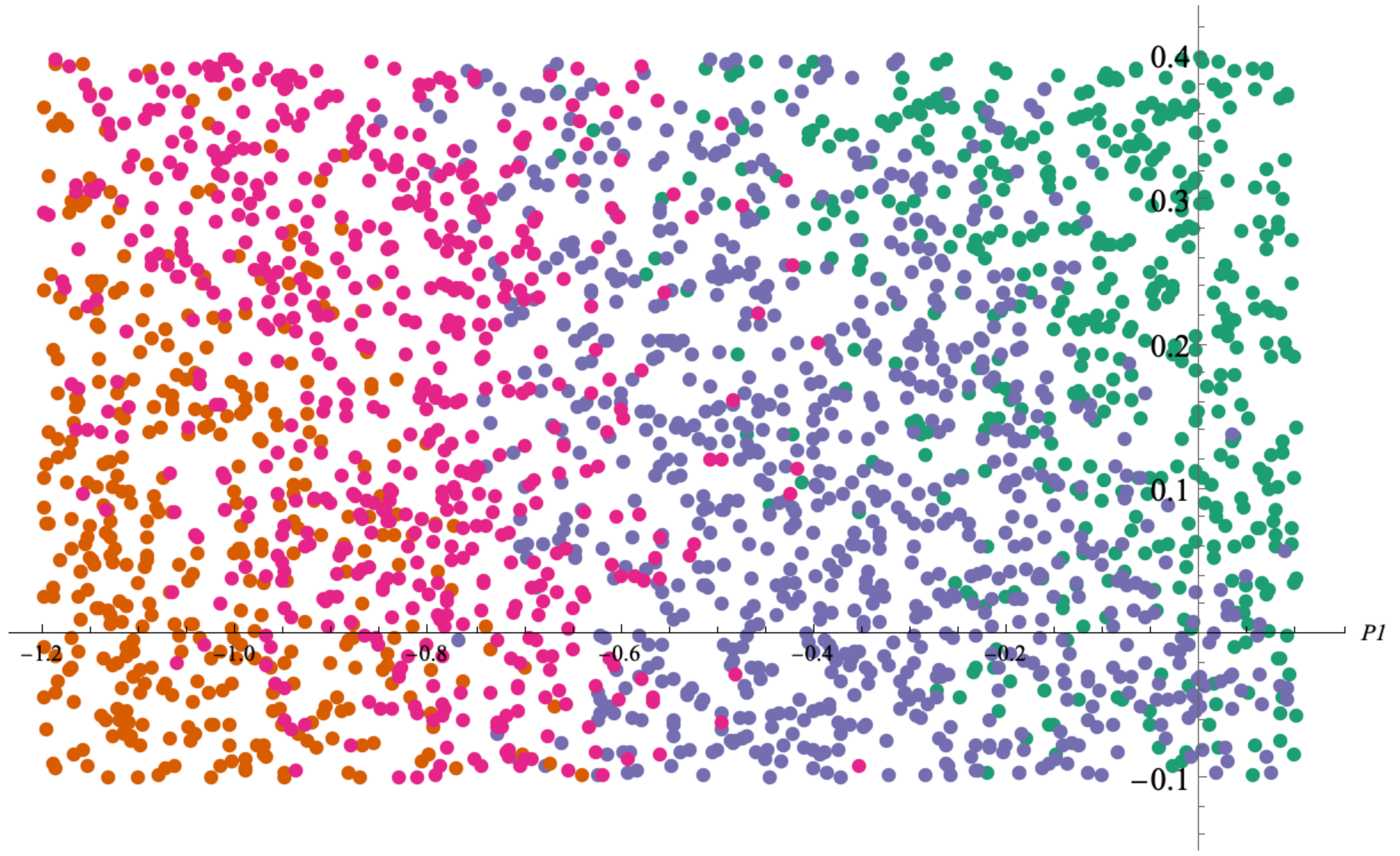


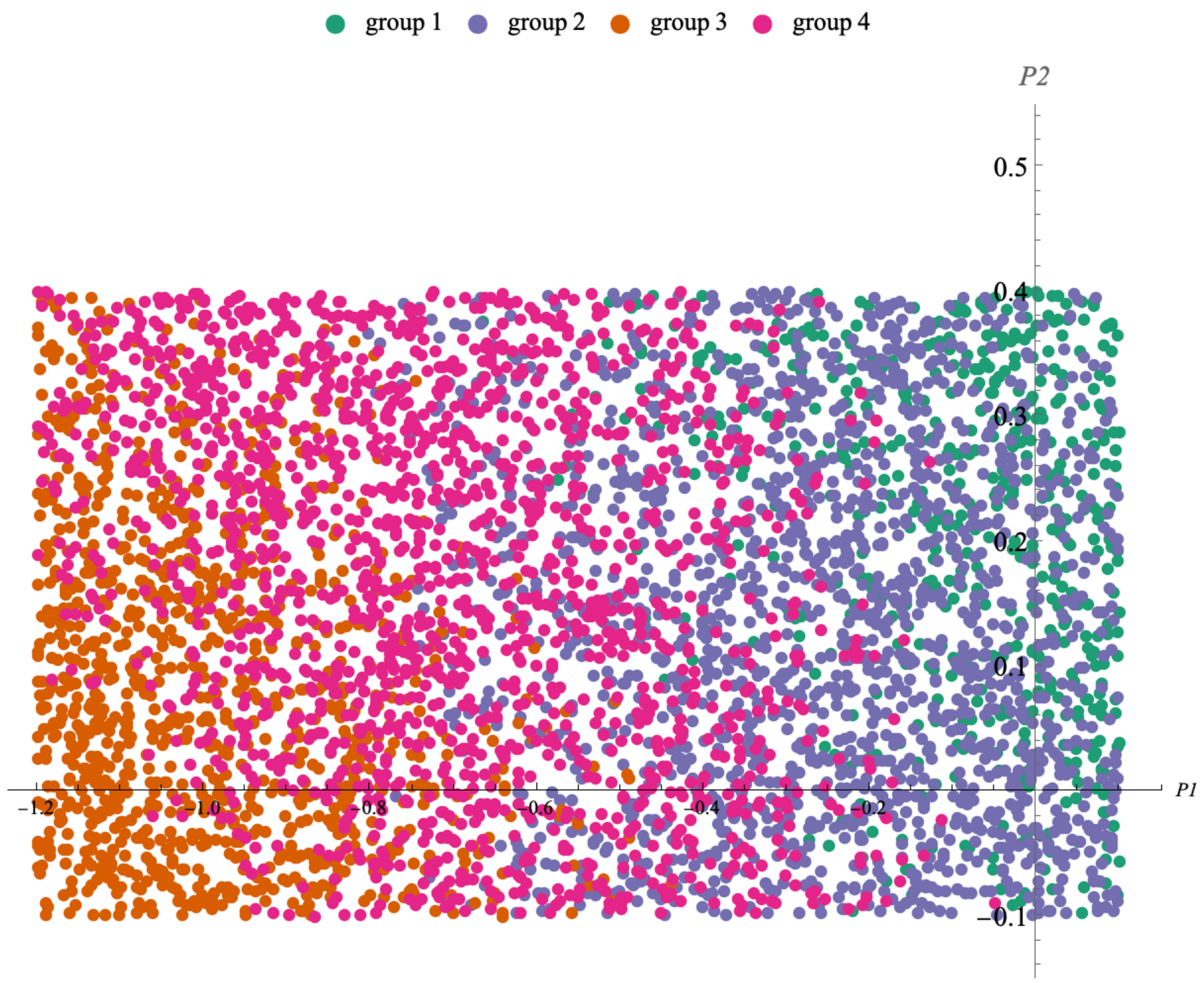
tour in observable space

$C_9 - C_{9'}$ varying slice width into projection



Thin slice



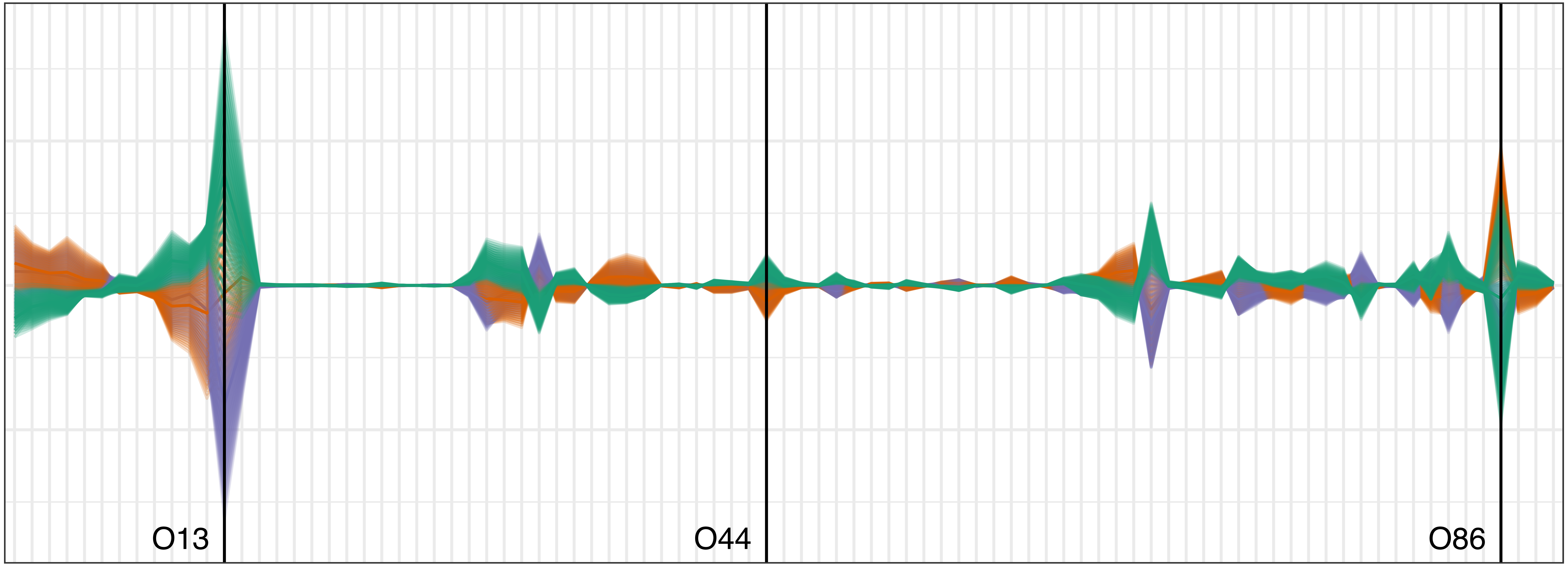


Projection

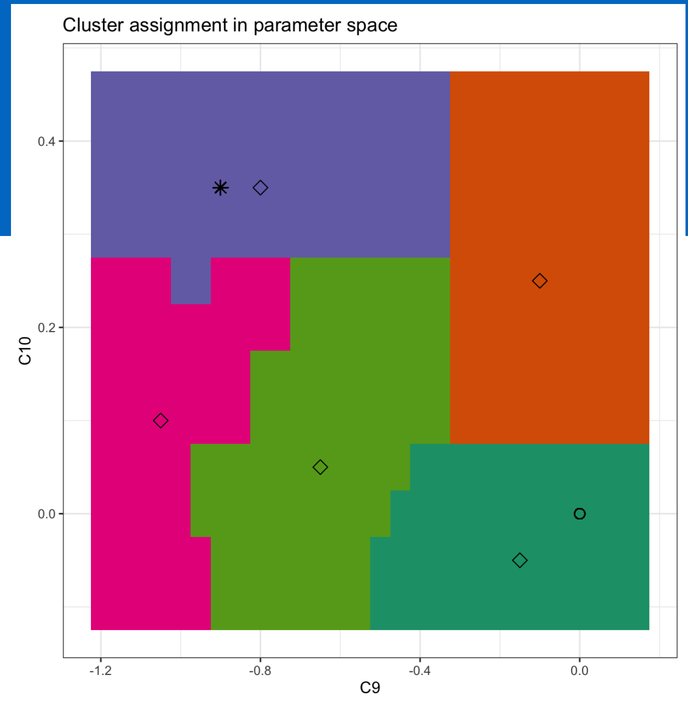
Include more observables

larger observable set

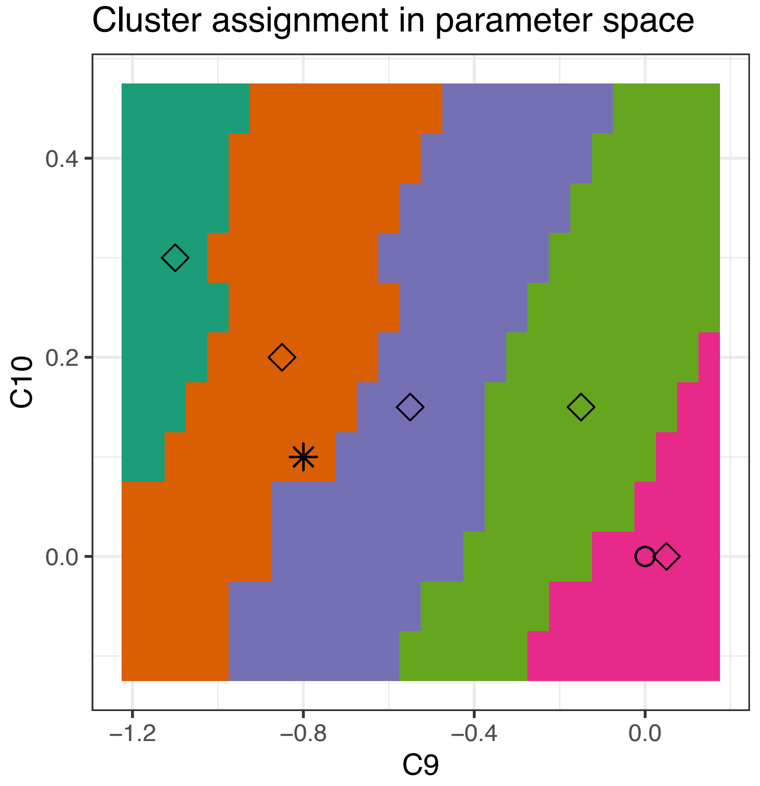
- as we know there are many more (~200) observables included in global fits.
- but not all have the same impact, here we look at a set of 89 and look first at the PC plot which suggests which ones merit closer scrutiny
  - 86 has the largest variation after R_K
  - 44 looks “small” but will come back to that



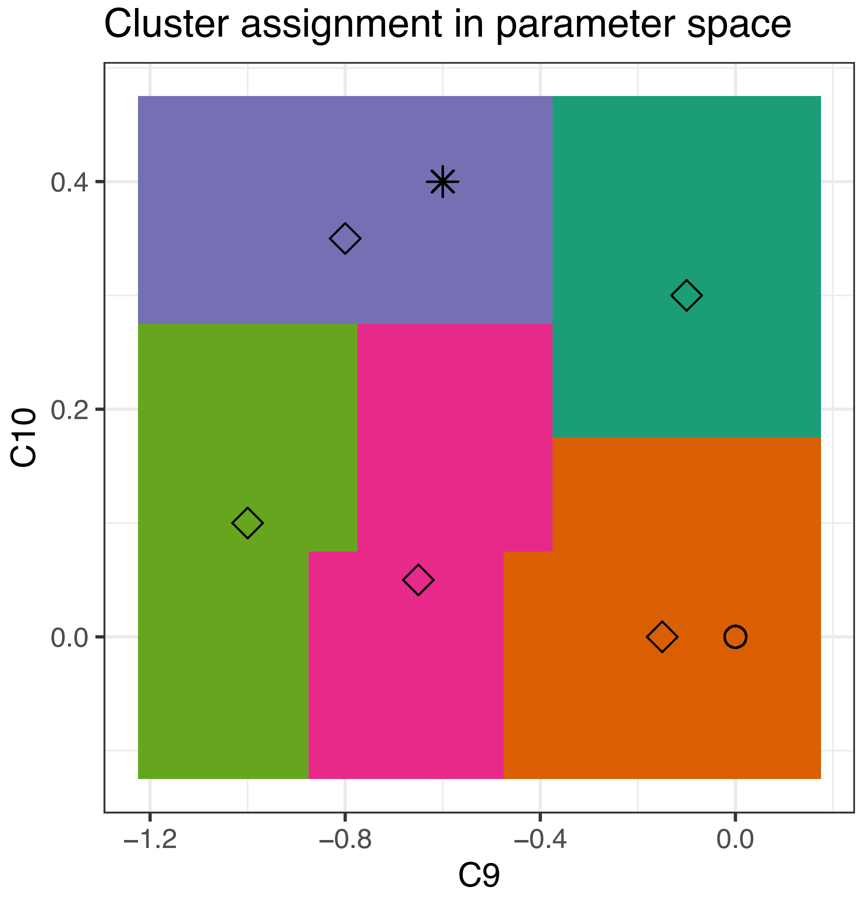
effect of (86) $\mathscr{B}(B_s \to \mu^+\mu^-)$ on original set

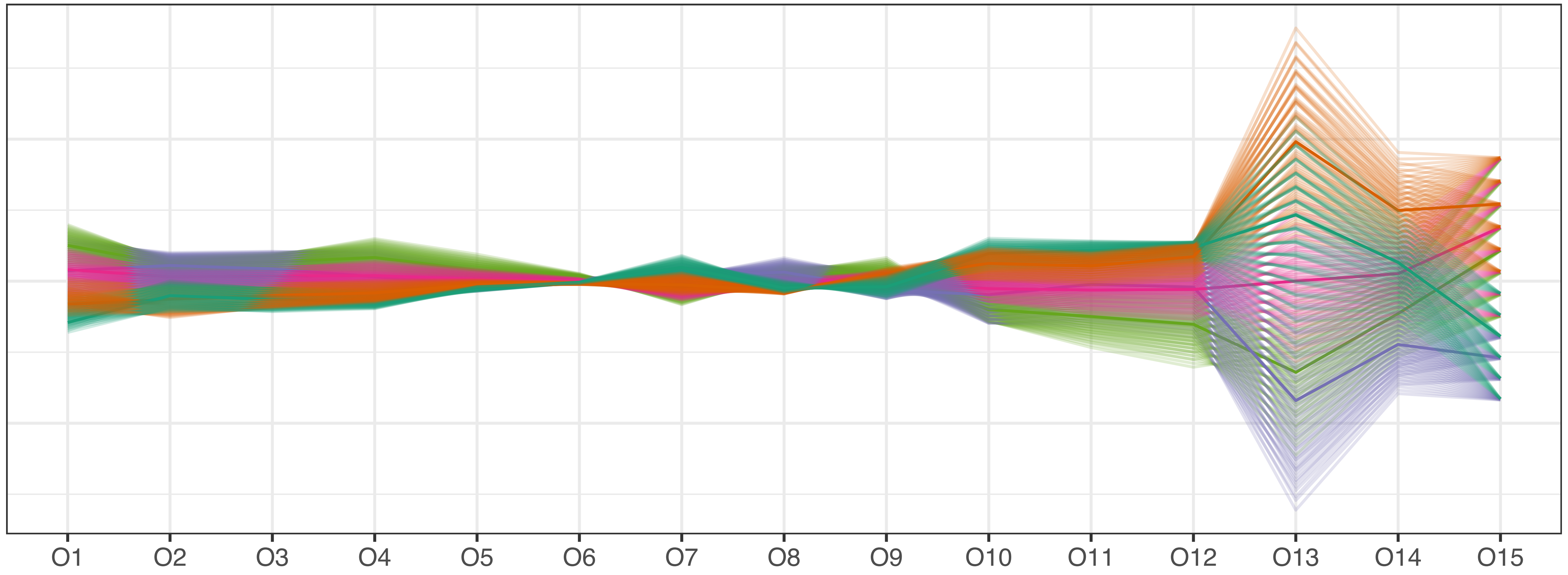


5 clusters with
all 89 observables



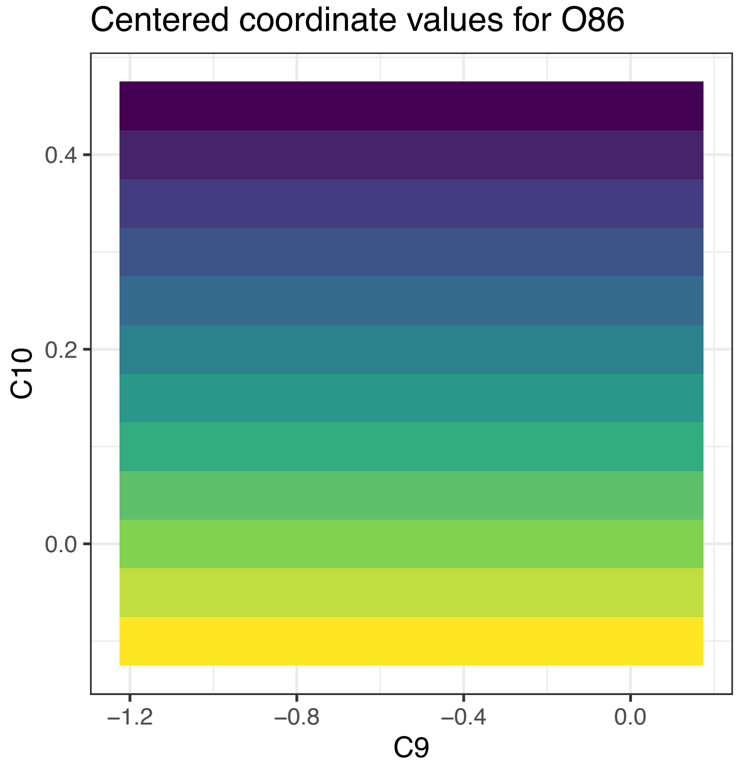
$\mathscr{B}(B_s \to \mu^+\mu^-) = (2.81^{+0.24}_{-0.22}) \times 10^{-9}$

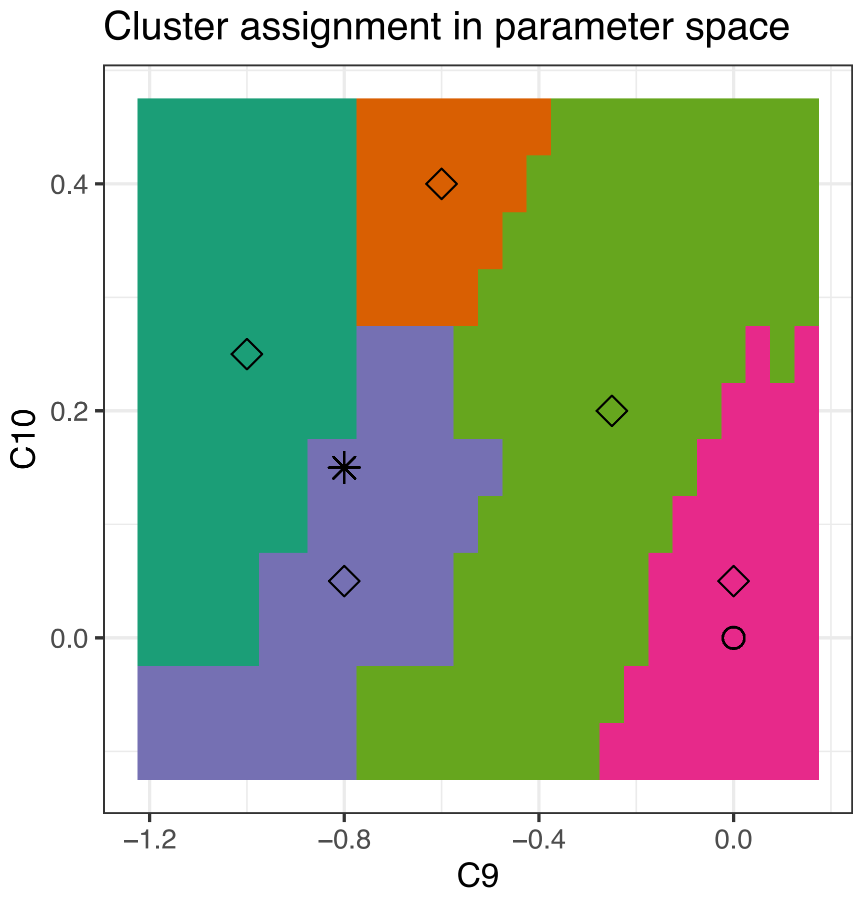


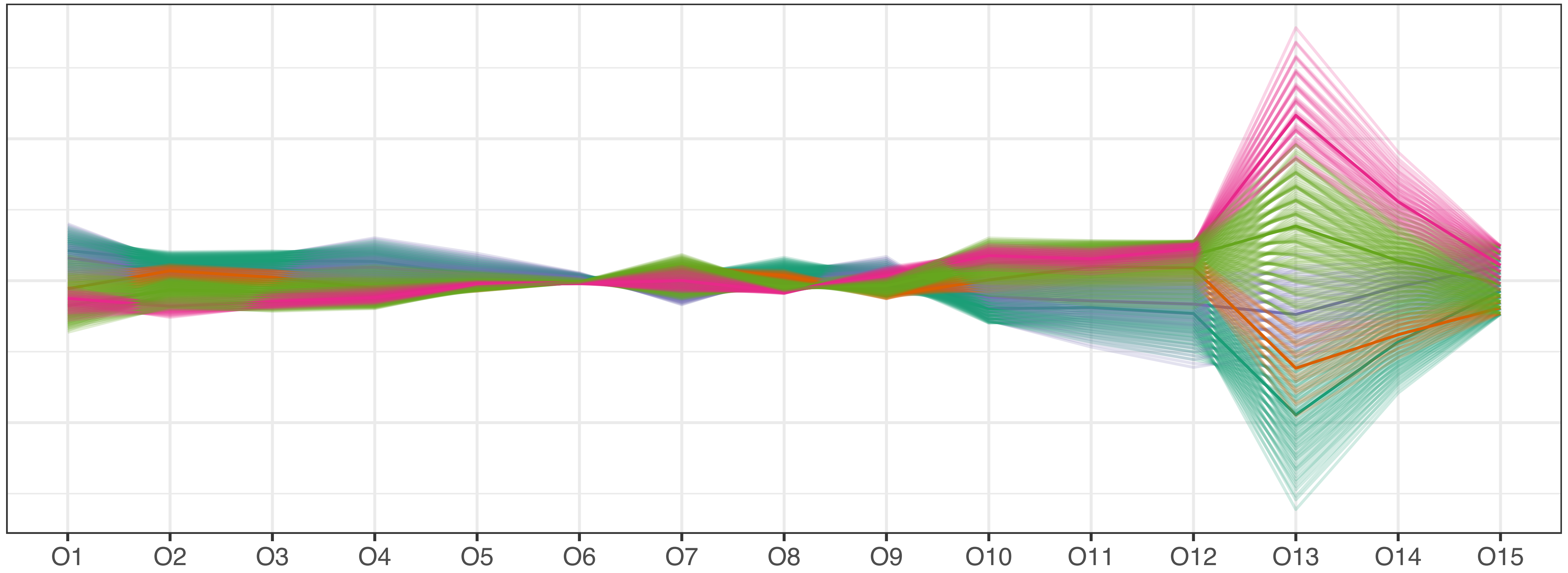


$O_{15} = \mathscr{B}(B_s \to \mu^+\mu^-)$

$\mathscr{B}(B_s \to \mu^+\mu^-) = (2.81 \pm 0.43) \times 10^{-9}$

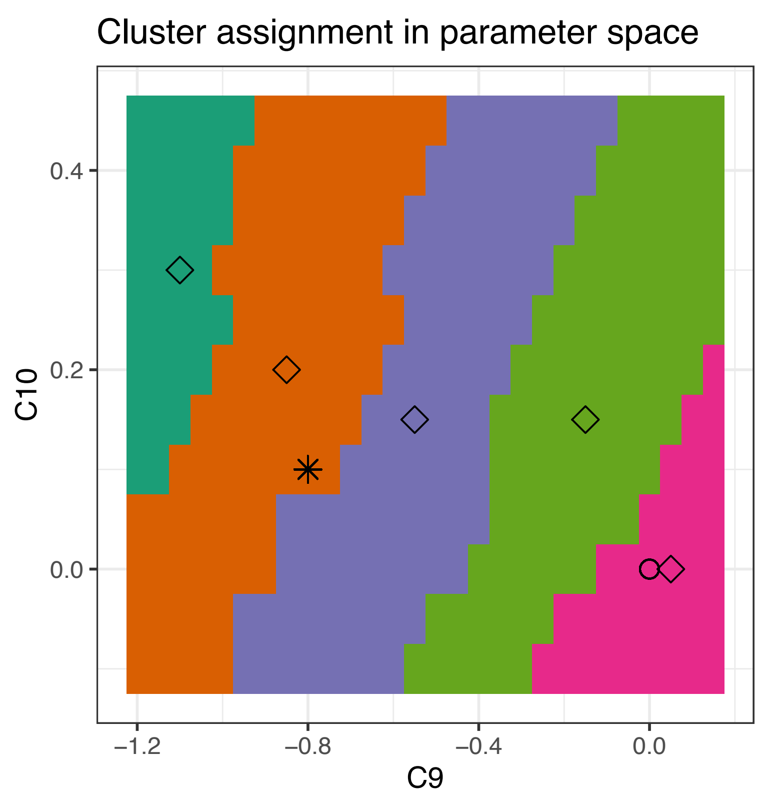




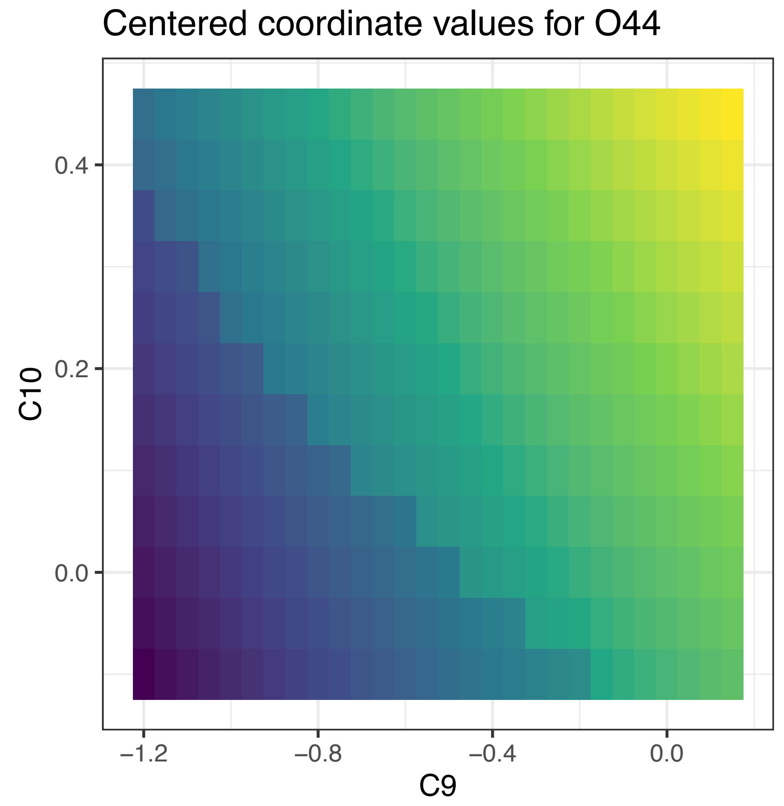


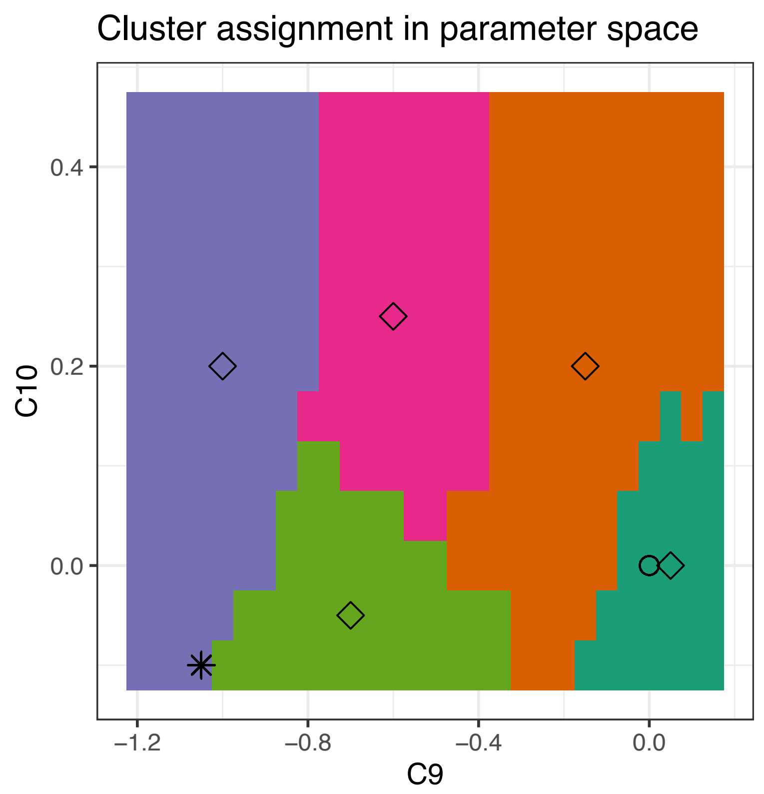
effect of (44) $P_4'[0.1 - 0.98]$ on original set

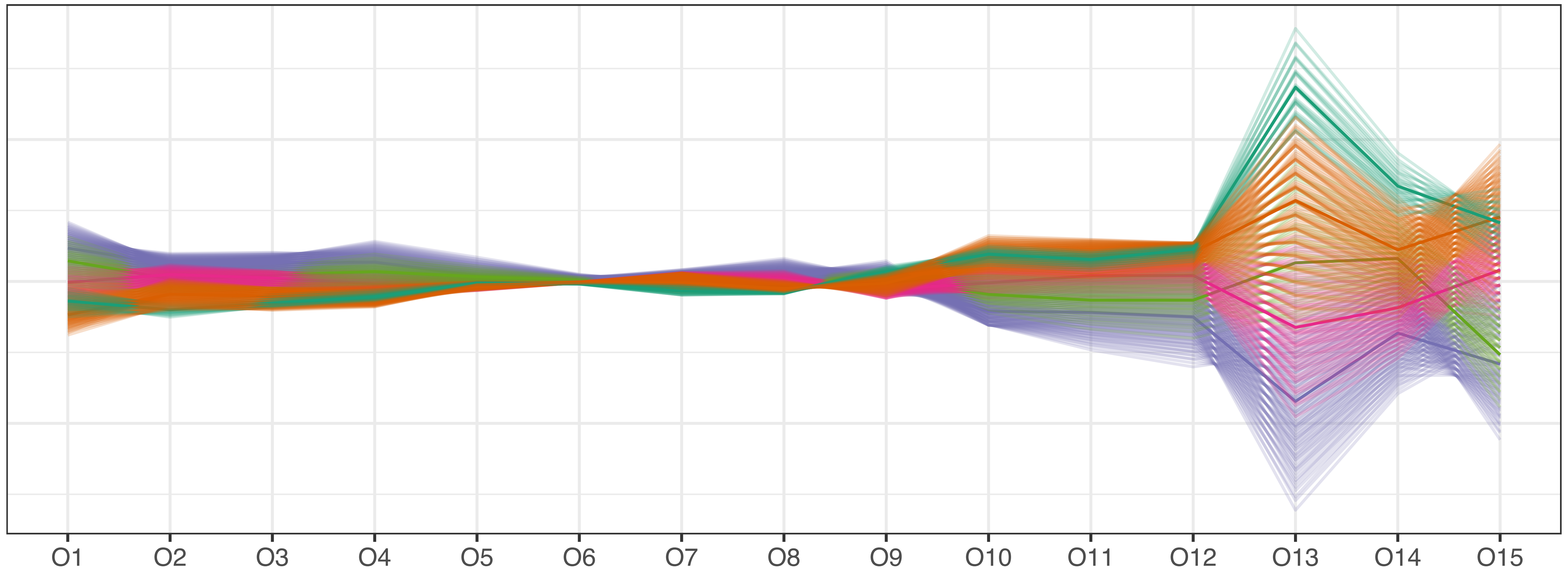
- currently $P_4'[0.1 - 0.98] = 0.135 \pm 0.118$ does not have a major impact
- if the error can be reduced by a factor of four then it can have a large impact



add 44 reduced error

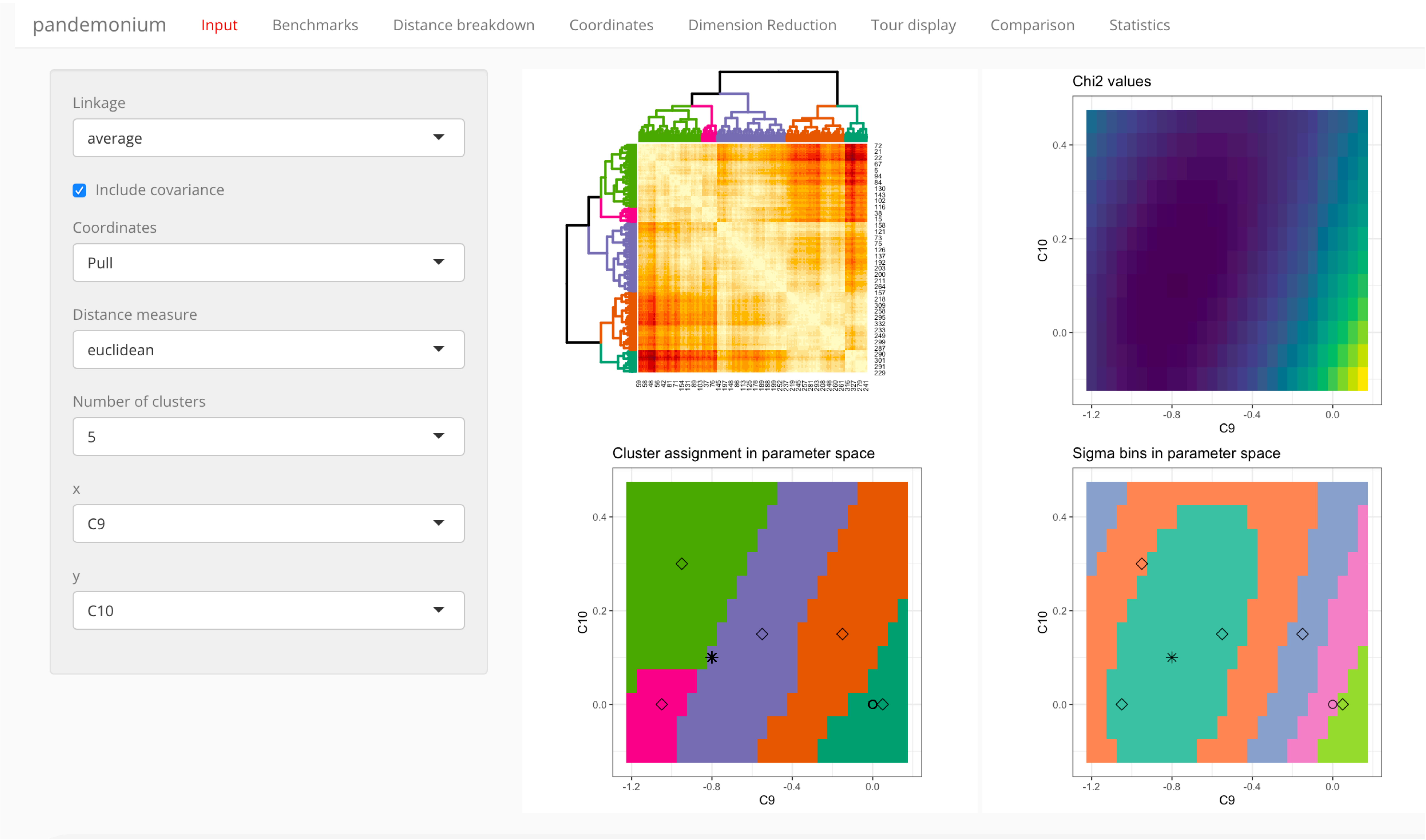






implementation by **Ursula Laa** available on GitHub as a Shiny app (an R package)

https://github. com/uschiLaa/pandemonium



Thanks