MONASH University

clustering and visualisation study of the B anomalies

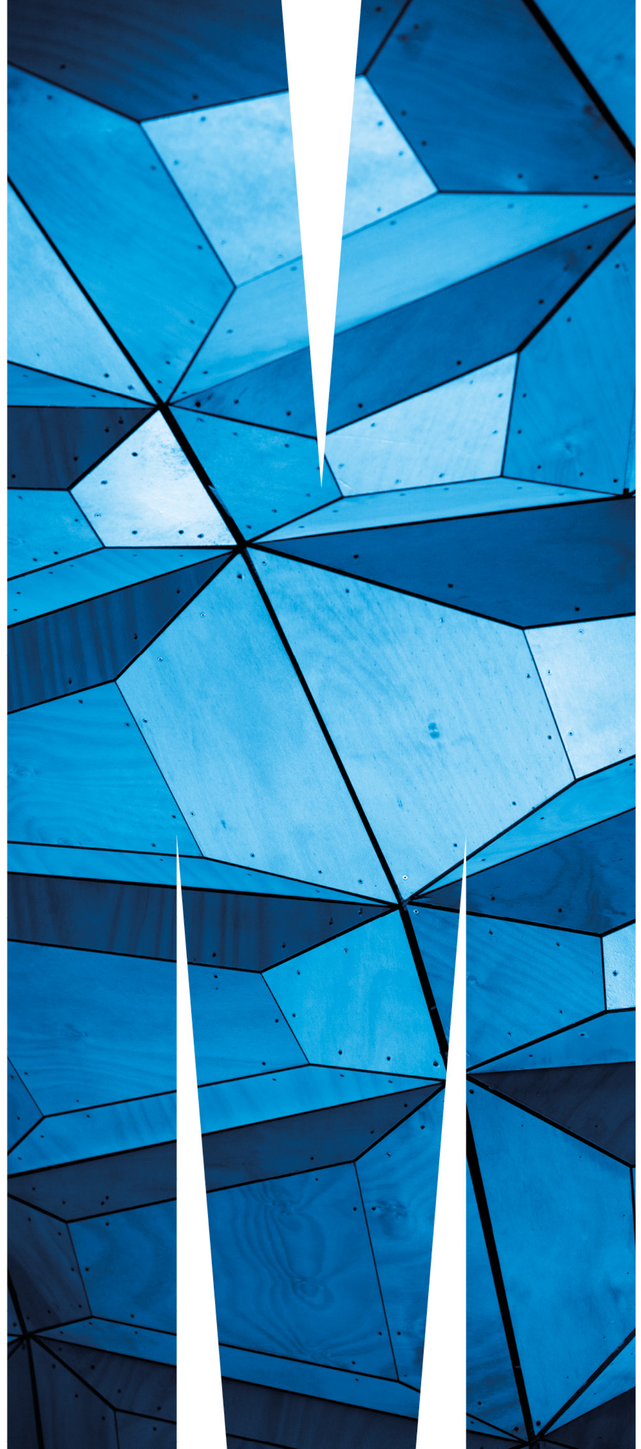
Anomalies 2021

German Valencia

based on: Ursula Laa (BOKU, Vienna) and G. V (Monash).:
Pandemonium: a clustering tool to partition parameter space
-- application to the B anomalies
2103.07937 [physics.data-an] and work in progress

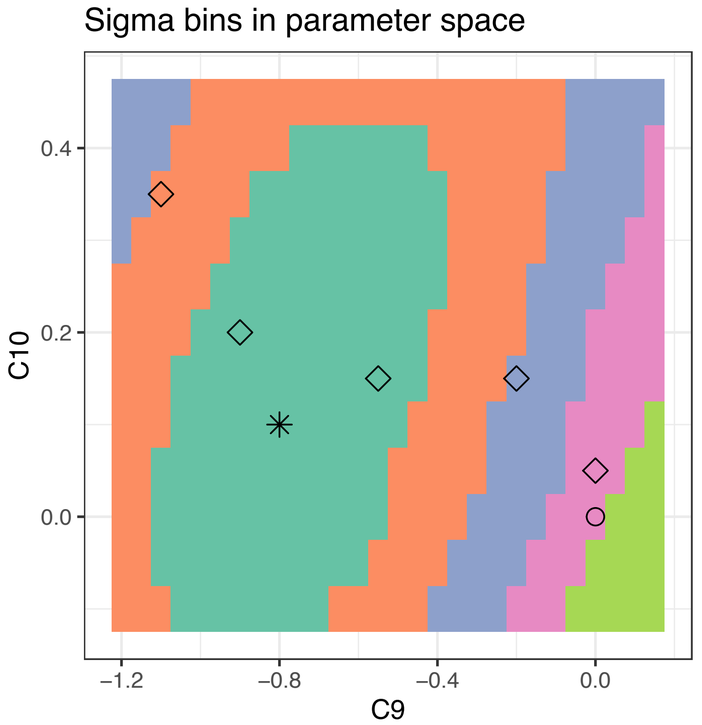
questions: German.Valencia@monash.edu
ursula.laa@boku.ac.at

GROUP OF EIGHT AUSTRALIA



global fits

- many existing global fits to large numbers of observables in $b \to s\ell^+\ell^-$ have shown a deviation from the SM at the ~5σ level
- global fits give results as:
  - goodness of fit
  - best fit parameters
  - confidence level intervals
- can be thought of as a clustering based on $\Delta\chi^2$
  - the clusters are the confidence intervals



beyond a global fit

- clustering
  - partition of parameter space into “clusters”
    - more than one way to do it emphasising different aspects
    - resolving power: how many different groups exist
  - find representative centroids for each cluster
    - these provide a small set of benchmark points for detailed studies
  - isolate trends and effects of subsets of observables
- visualisation
  - view of the observable space (typically high dimensional)
  - relative importance of observables: prioritising for further study
  - collective observable dependence on parameters
  - 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 suggest some parameters are more important than others:

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

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

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

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

- allow NP for muons only

_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

- pull- residual analysis ranks observables
  Eur.Phys.J.C 79 (2019) 6, 462
- ★ 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 lies 3.7σ away from the SM and is $(C_9, C_{10}) = (-0.8, 0.1)$

distance between models

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

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

- the "origin' O_i is any reference point: experiment E_i, SM, ⋯
- 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

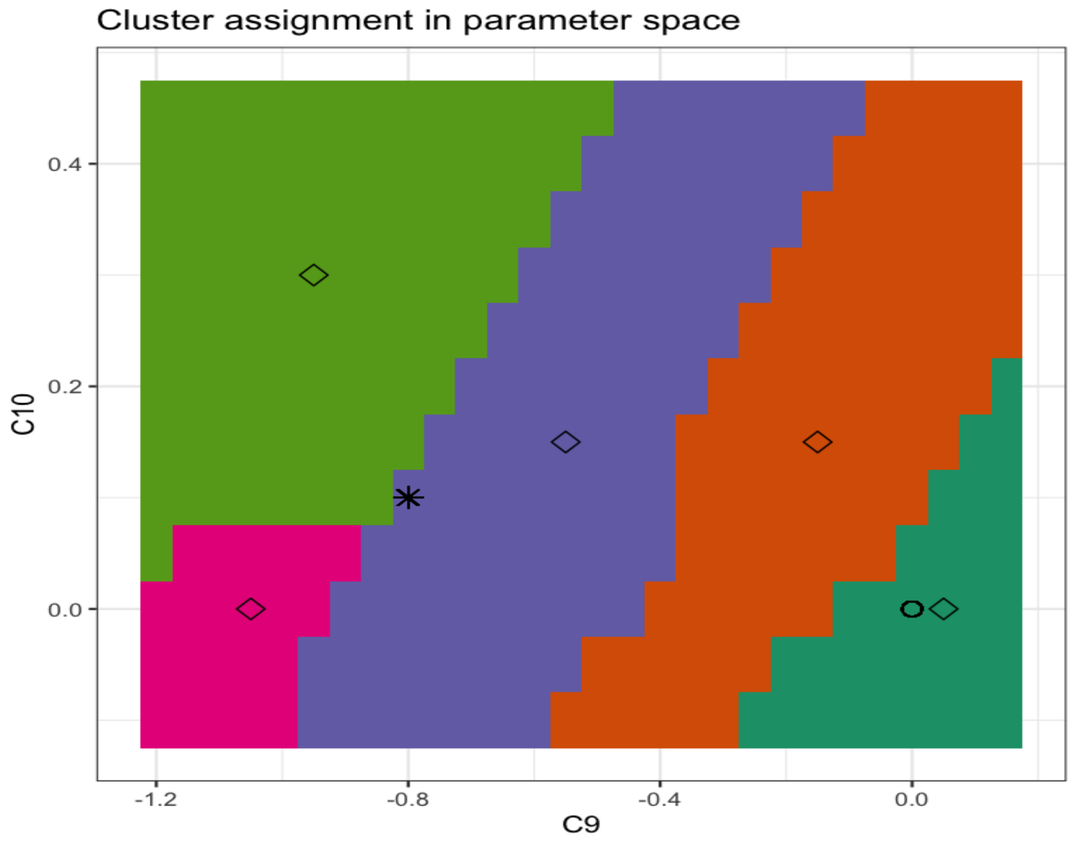
$$\chi^2_k = \sum_{i,j} [X_{ki} - E_i](\Sigma^{exp} + \Sigma^{th})^{-1}_{ij}[X_{kj} - E_j] = \sum_i Y^2_{ki}$$

- analogously, the (Mahalanobis) distance between models

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

- the last equality follows if Σ does not depend on the model
- clustering takes into account all inter-point distances and not just distance to a reference point
- this distance it can be interpreted as a $\Delta\chi^2$

centroid and radius



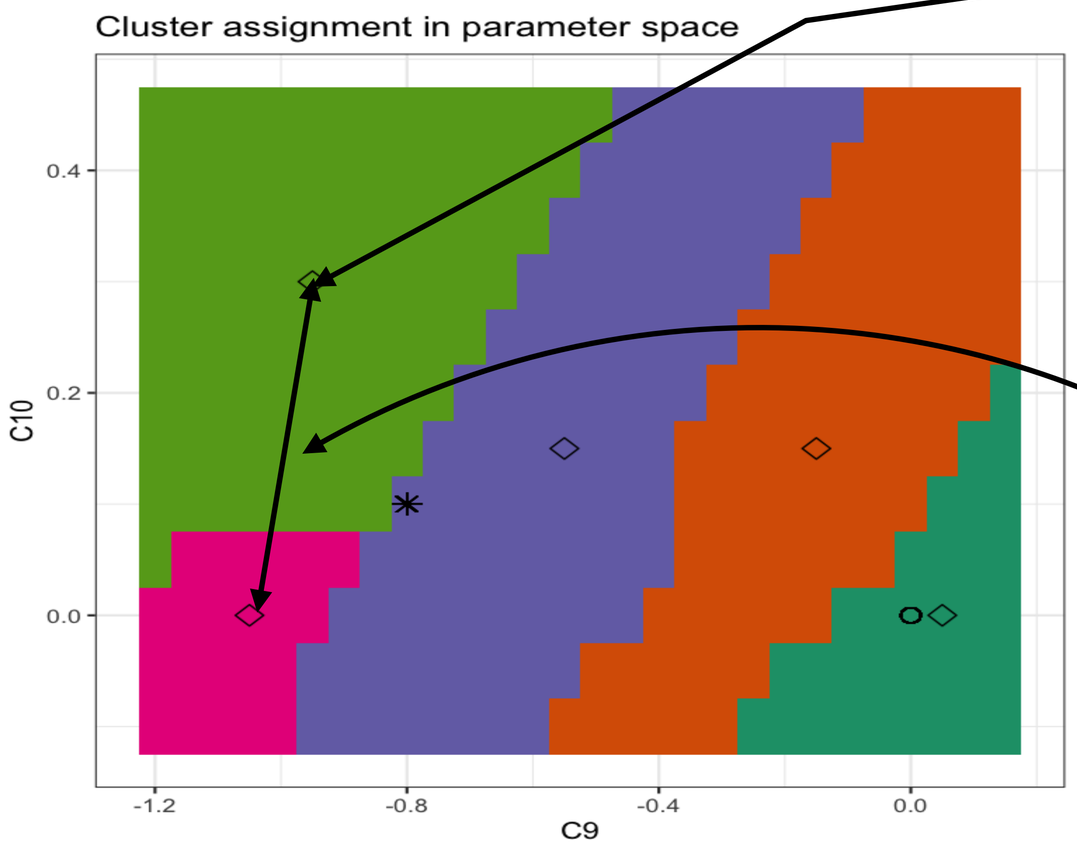
- want centroids to be representative of their cluster
- centroids will serve as benchmark points
- The centroid c_j is the member of the cluster C_j which minimises

$$f(c, C_j) = \sum_{x_i \in C_j} d(c, x_i)^2$$

- The radius of the cluster is

$$r_j = \max_{x_i \in C_j} d(c_j, x_i)$$

“one sigma” clusters



- 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

- the distance between any two centroids would be at least 1σ: $\Delta\chi^2 \geq 2.3$
- 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

now cluster the 14 observables

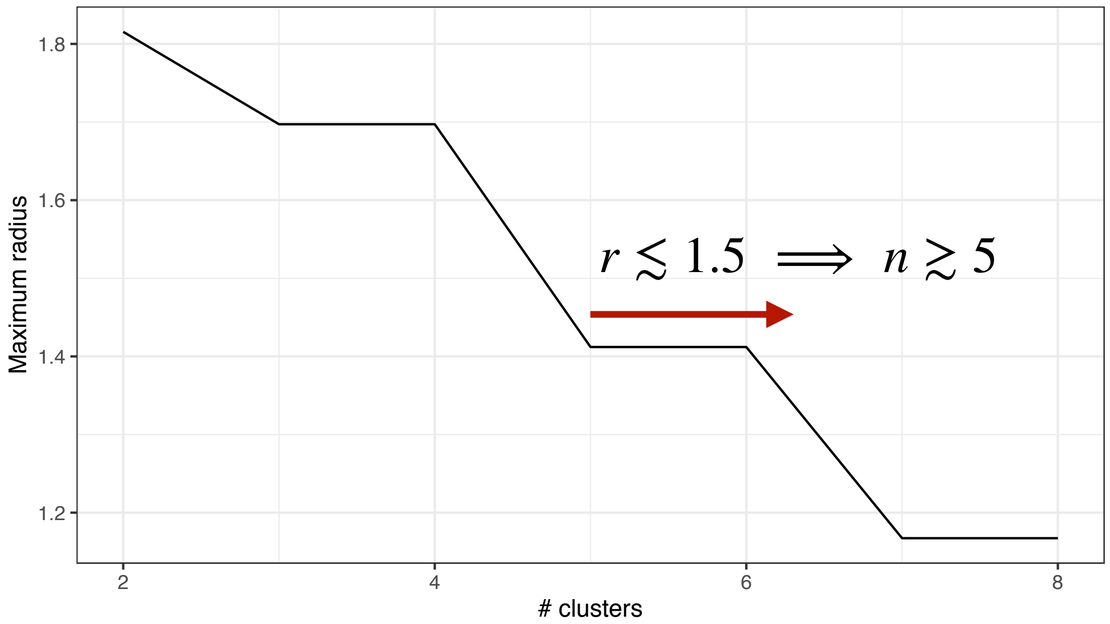
- generate models (sets of 14 predictions) on a uniform grid of parameters C_9^μ, C_{10}^μ
  - Uniform grid is needed for current visualisation tool
  - All data shown was generated with `flavio` (D. Straub, "flavio: arXiv: 1810.08132)
- the grid includes the BF and the SM and points in the region between them
- in what follows we use Euclidean distance on the pulls with Ward linkage unless otherwise specified

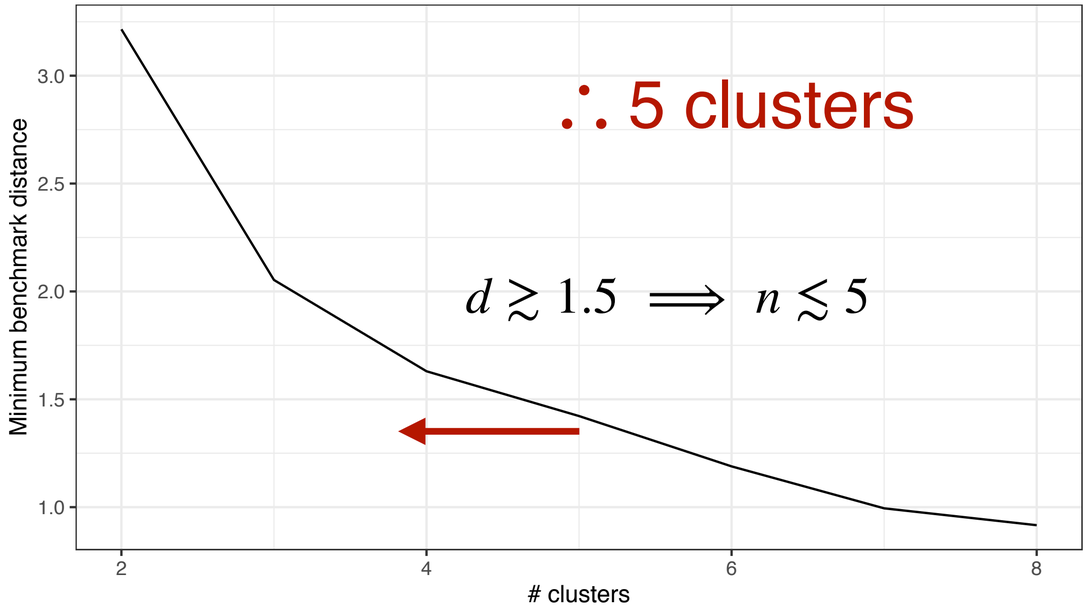
how many clusters? -euclidean distance

- square of euclidean distance on pulls $\approx \Delta\chi^2$

- 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

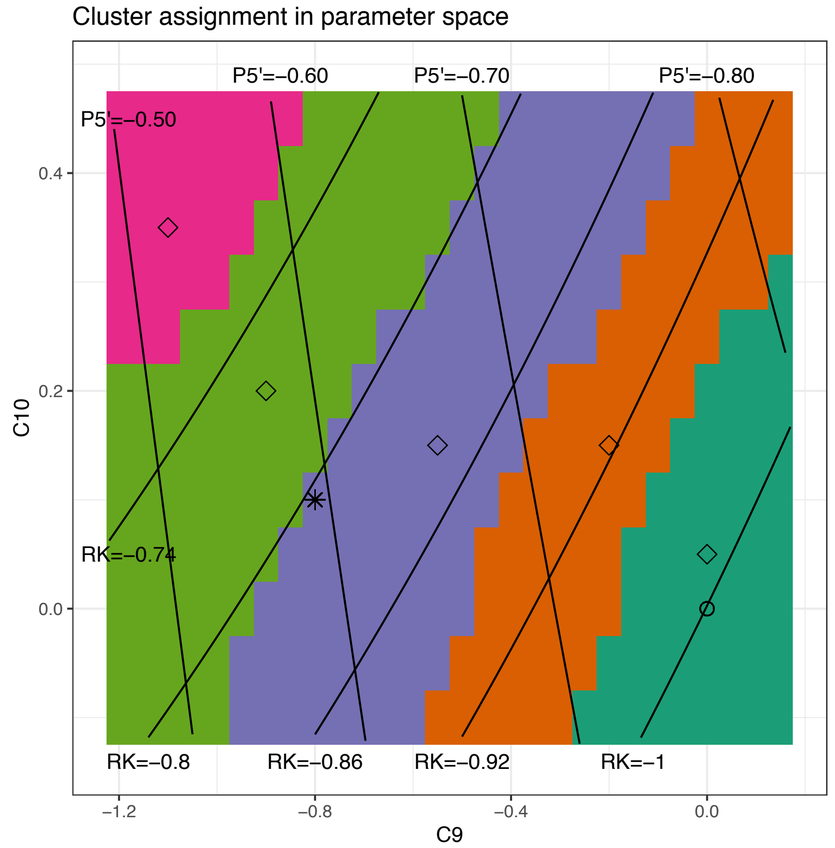




resolving power

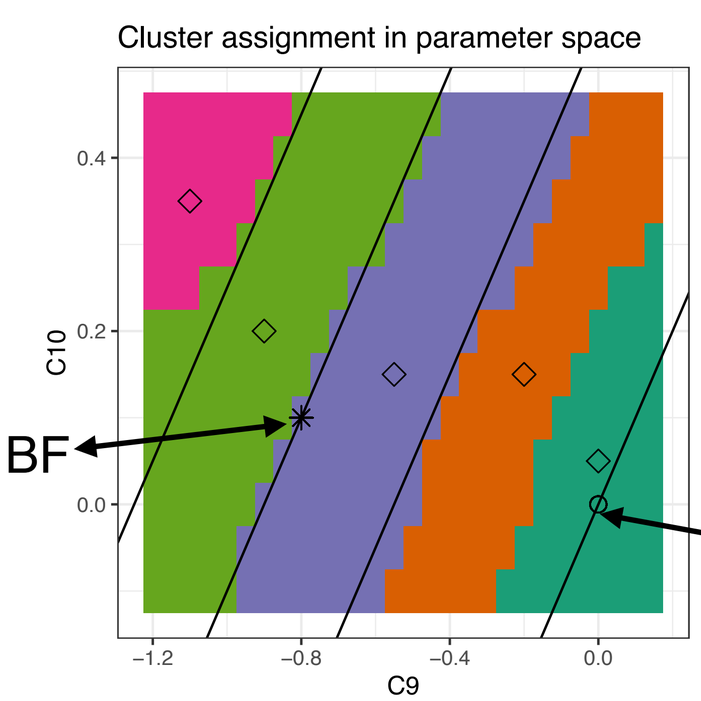
- number of partitions of parameter space with a set of measurements
- resolving power of these 14 observables with current accuracy is five clusters
- depends on
  - parameter space volume
  - range of predictions for a given observable over that region of parameter space
  - experimental and theoretical 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

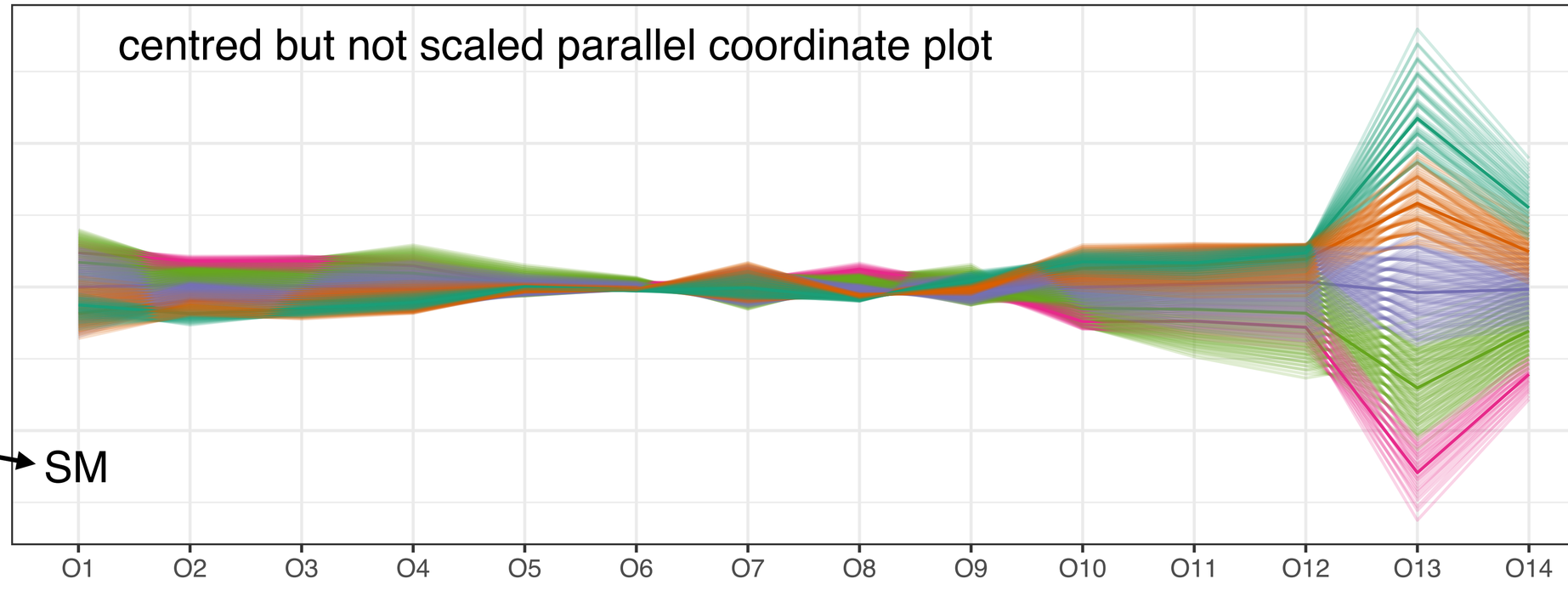
functional dependence of observables in parameter space



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

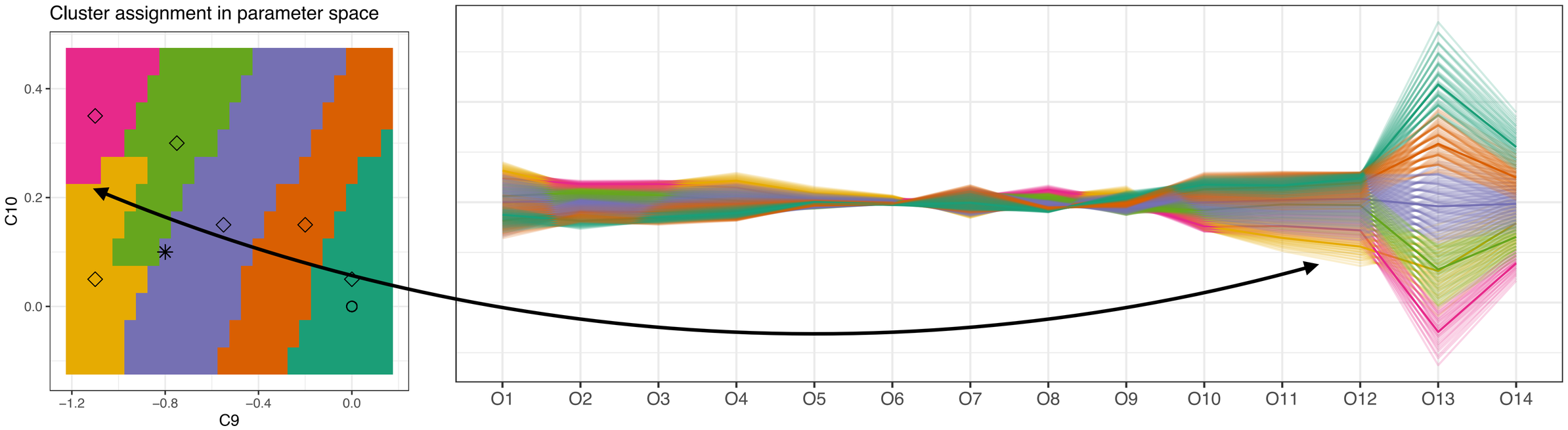
parameter vs observable space





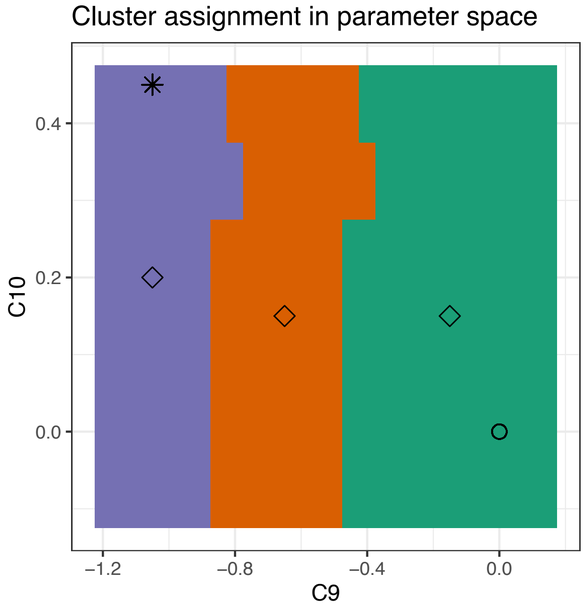
- Map clusters in two spaces: clusters are clearly separated by observables 2, 8, 13 and 14
- 13 (R_K) plays a dominant role because its predicted value (in pull units) varies the most across the parameter region
- If an operator is dominant, as R_K is in this case, it completely determines the inter cluster boundary shape.

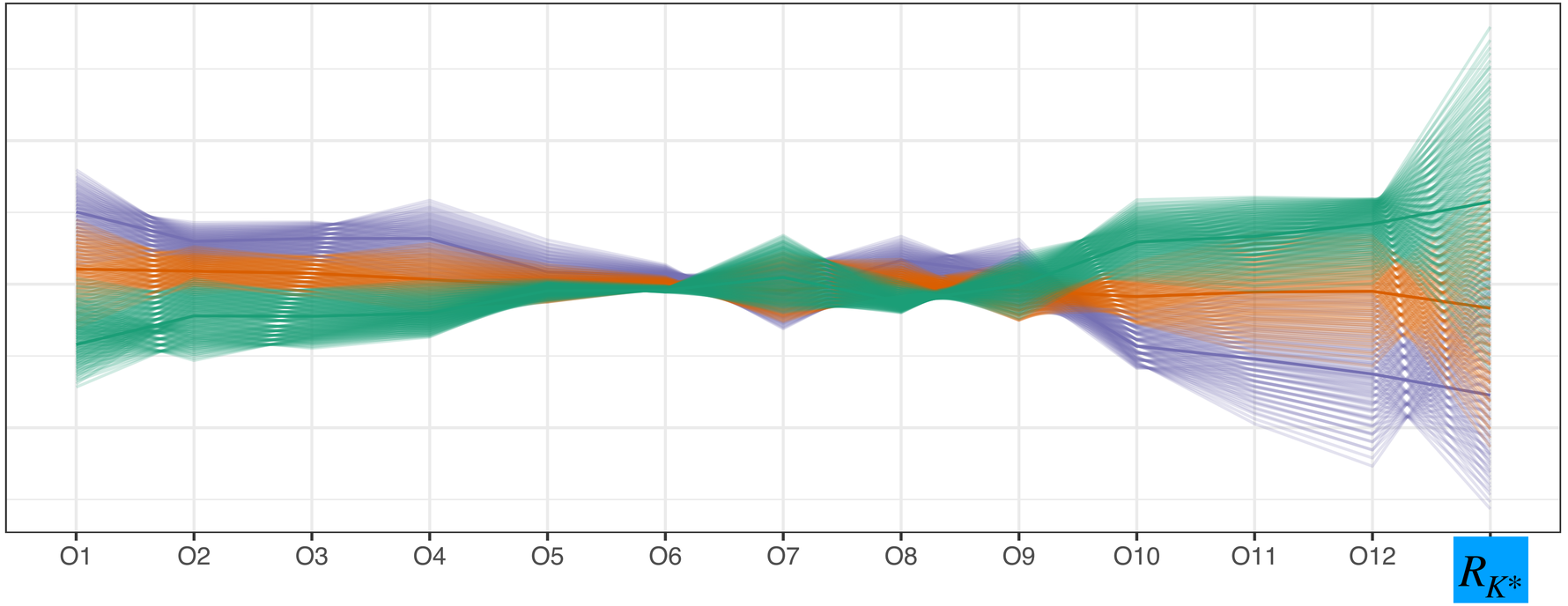
sub-leading effects



- 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

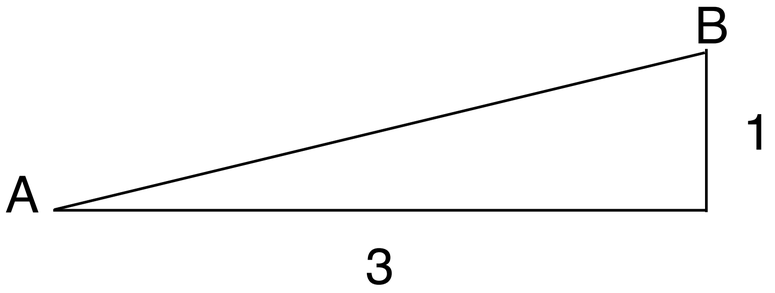
remove R_K





- resolving power reduced to about 3 clusters
- no one observable 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

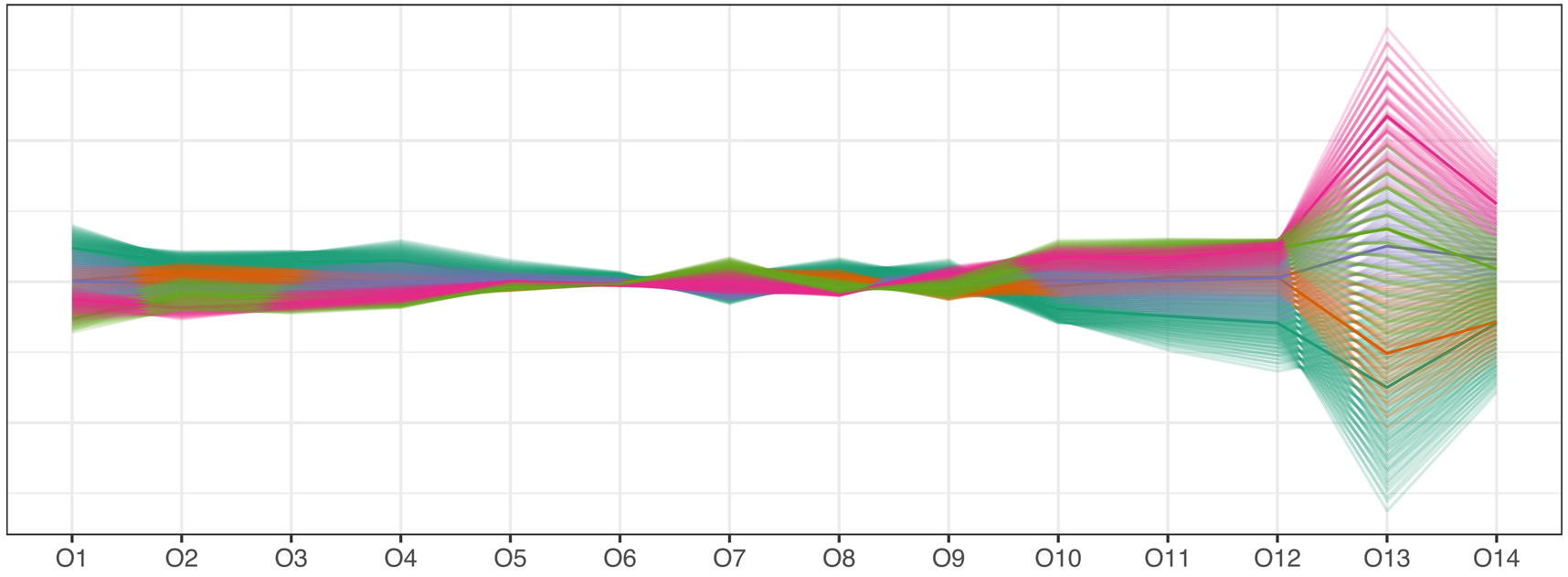
increasing the importance of sub-dominant observables without removing any

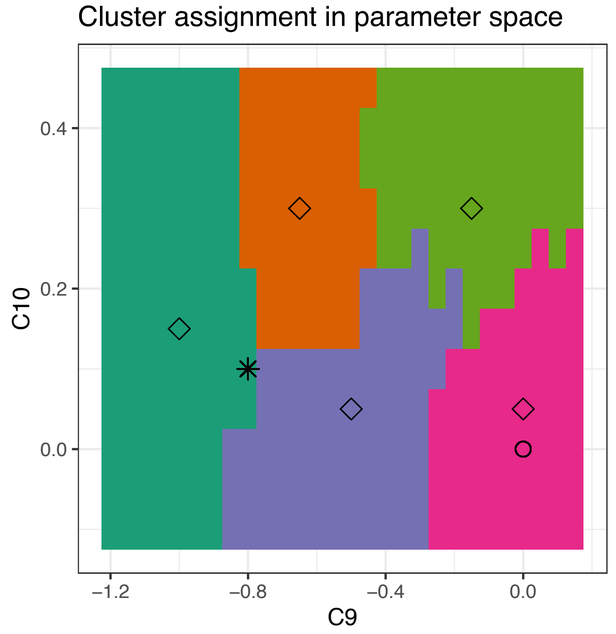


$$d_{\text{Manhattan}}(A, B) = 4$$

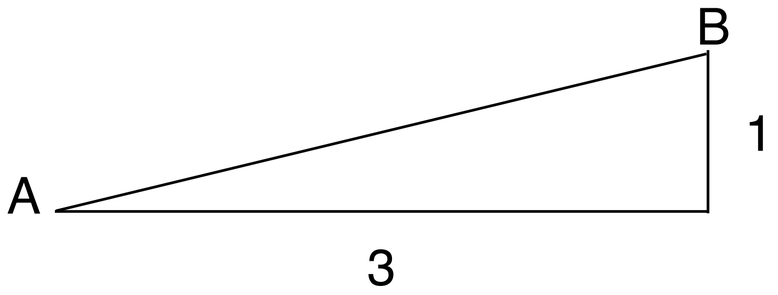
$$d_{\text{Euclidean}}(A, B) = 3.16$$

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





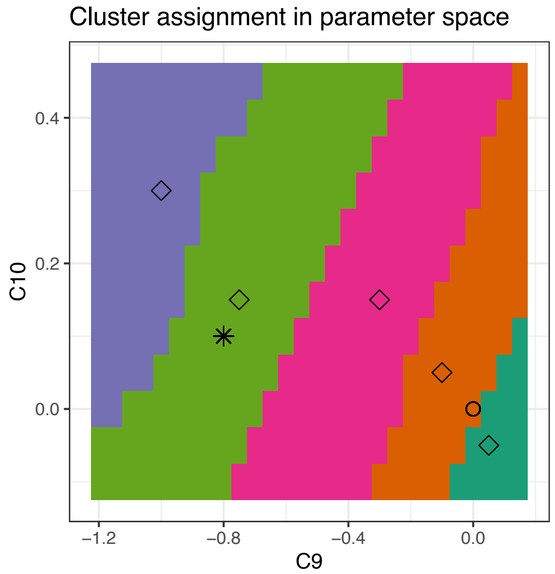
increasing the importance of dominant observables

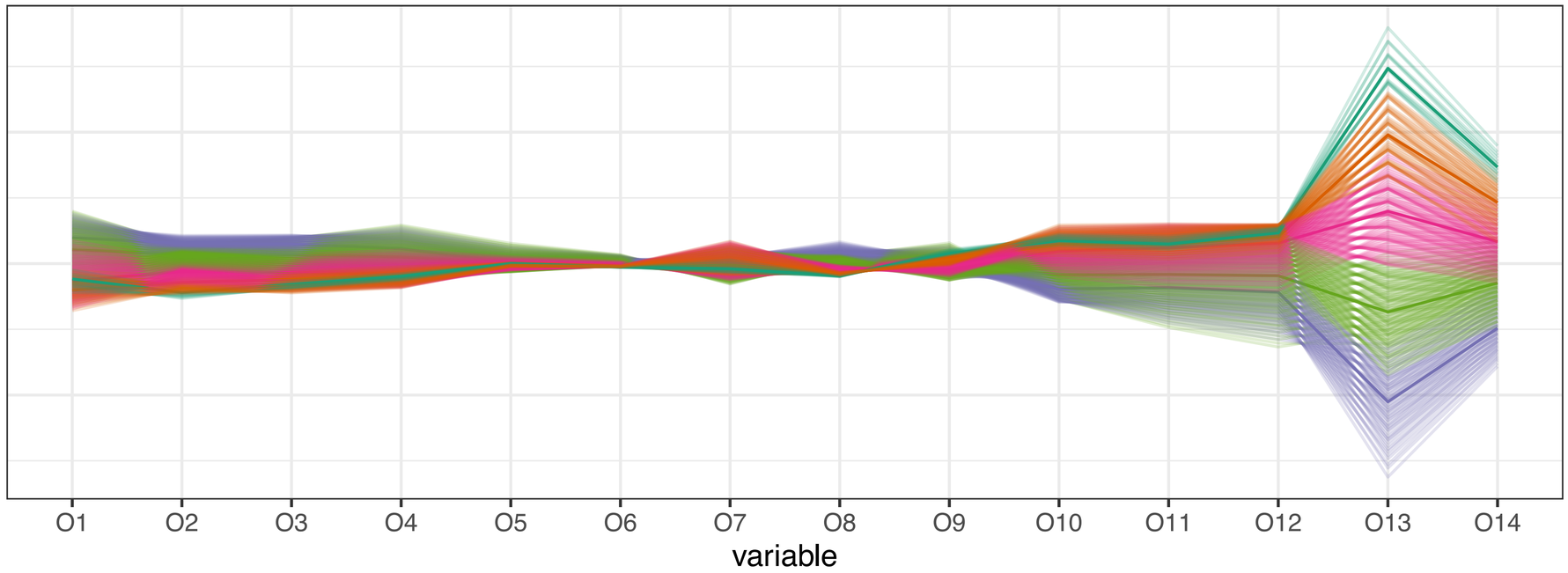


$$d_{\text{Chebyshev}}(A, B) = 3$$

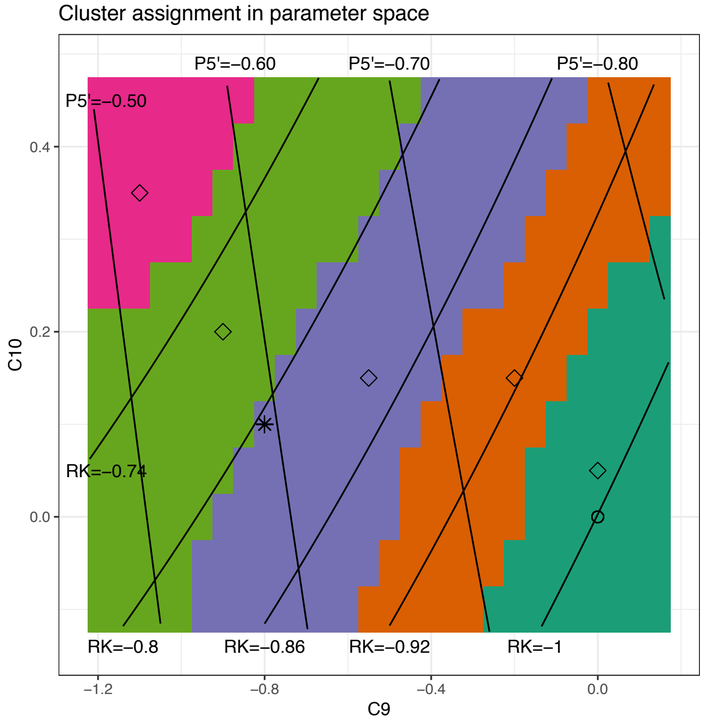
$$d_{\text{Euclidean}}(A, B) = 3.16$$

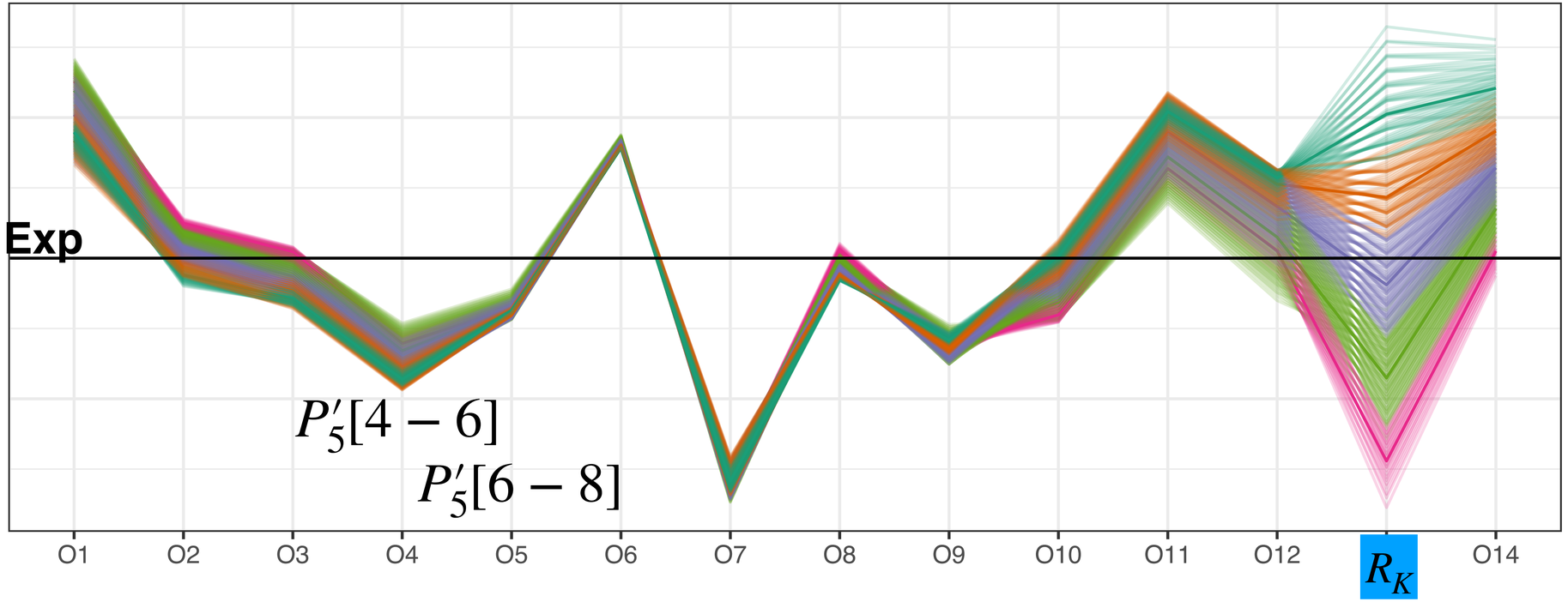
- five clusters with maximum distance and complete linkage





internal tensions in the fit

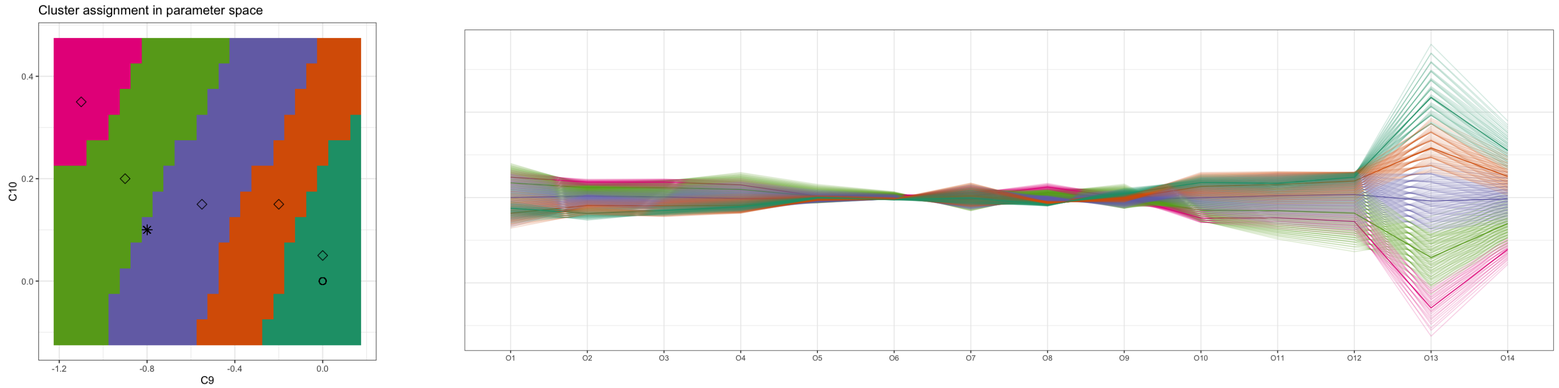




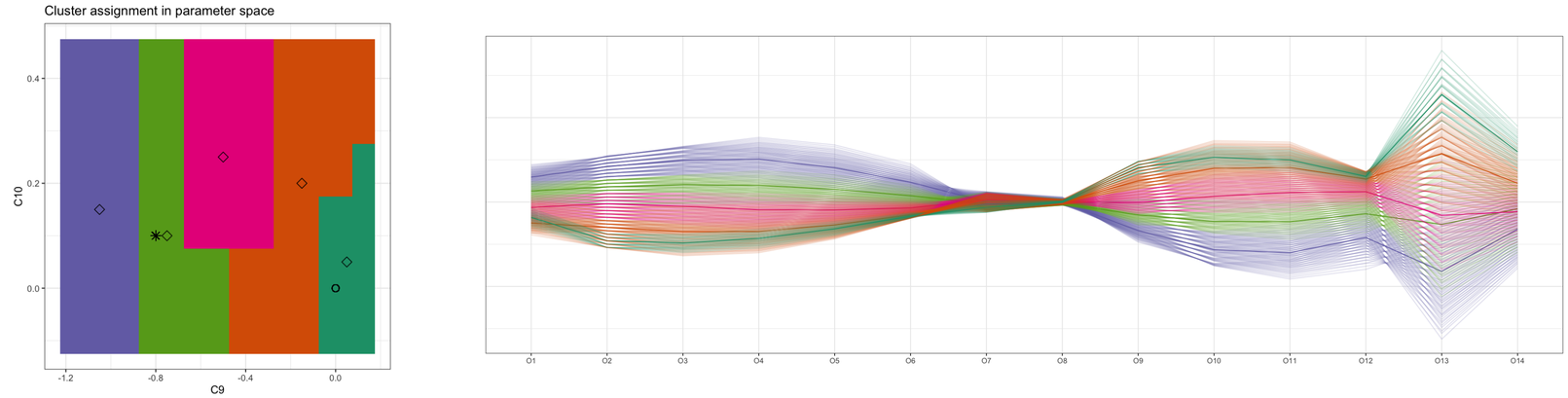
- BF on the boundary between light green and purple clusters
- $P_5'[4-6],\ P_5'[6-8]$ (largest discrepancy between SM and BF in P_5') prefer the light green: larger negative C_9 (recall the experimental value $P_5'[4-6] = -0.39 \pm 0.11$)
- $R_K = 0.86 \pm 0.06$ prefers the purple cluster (but not R_{K*})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

impact of correlations

- Can visualise the impact of correlations. In this example, it is known that the importance of P_5' is increased if the correlations are not included. This is reflected in a change from the previous result

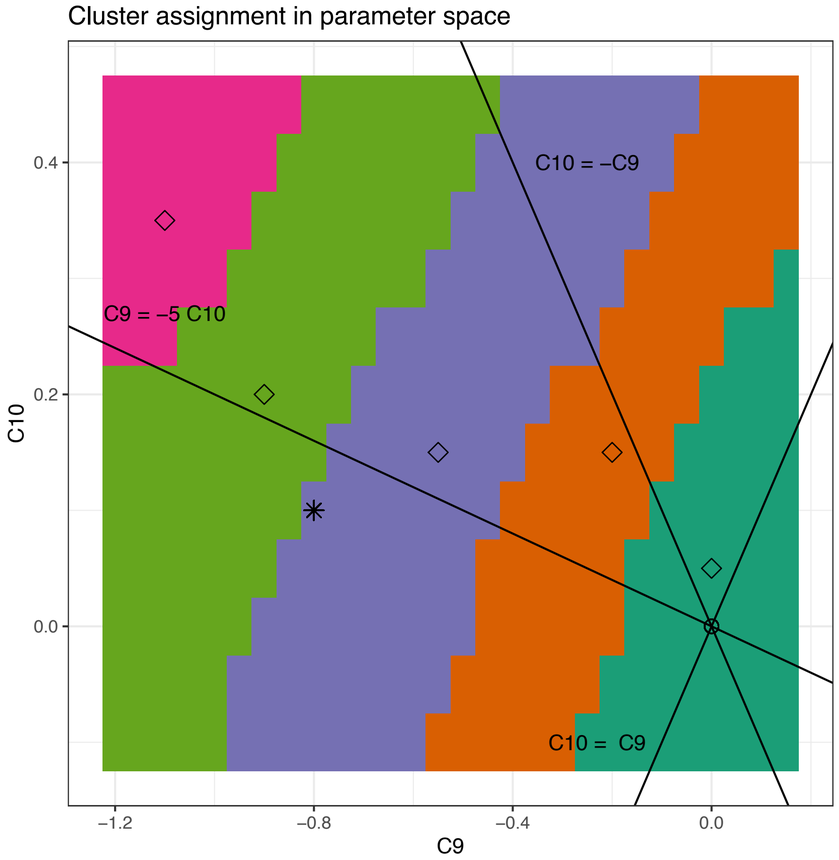


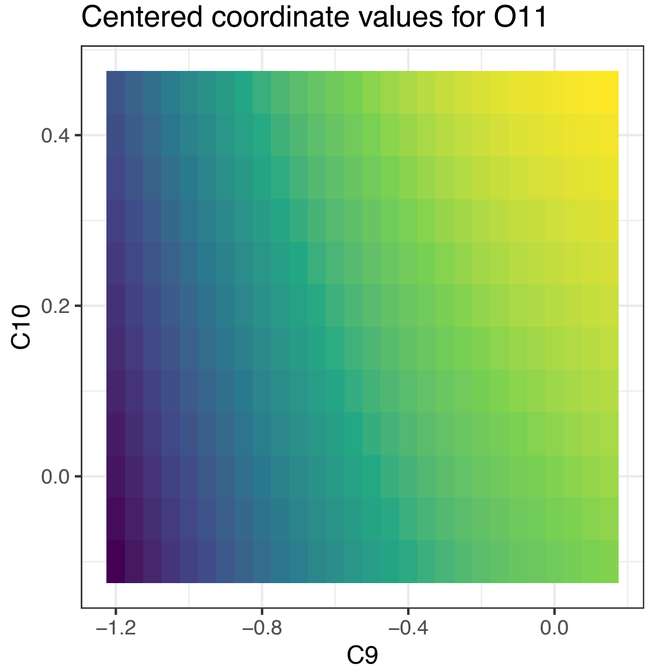
- to this one: (notice in parameter space boundaries closer to lines of constant P_5', in observable space larger range spanned by P_5' coordinates)



sensitivity along certain directions

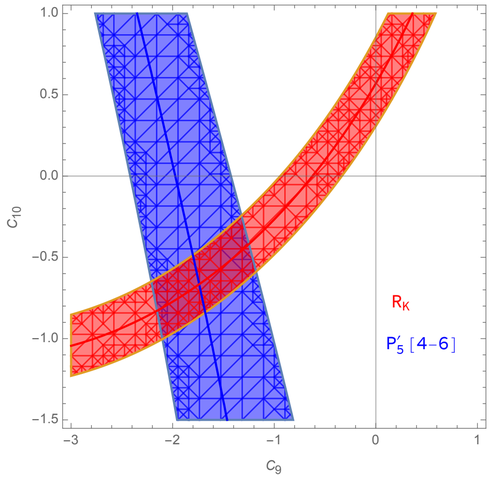
- most sensitive to a direction along $C_{10} \approx 0.2\ C_9$
- almost no sensitivity along $C_{10} = C_9$
- these patterns agree with what is seen in 2d global fits
- With this set, sensitivity along $C_{10} = C_9$ can be increased with a more precise measurement of $O_{11} = P_2[6-8]$



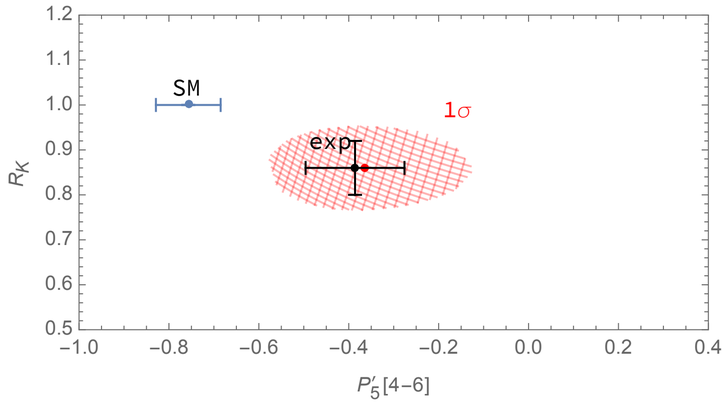


overall view in observable space

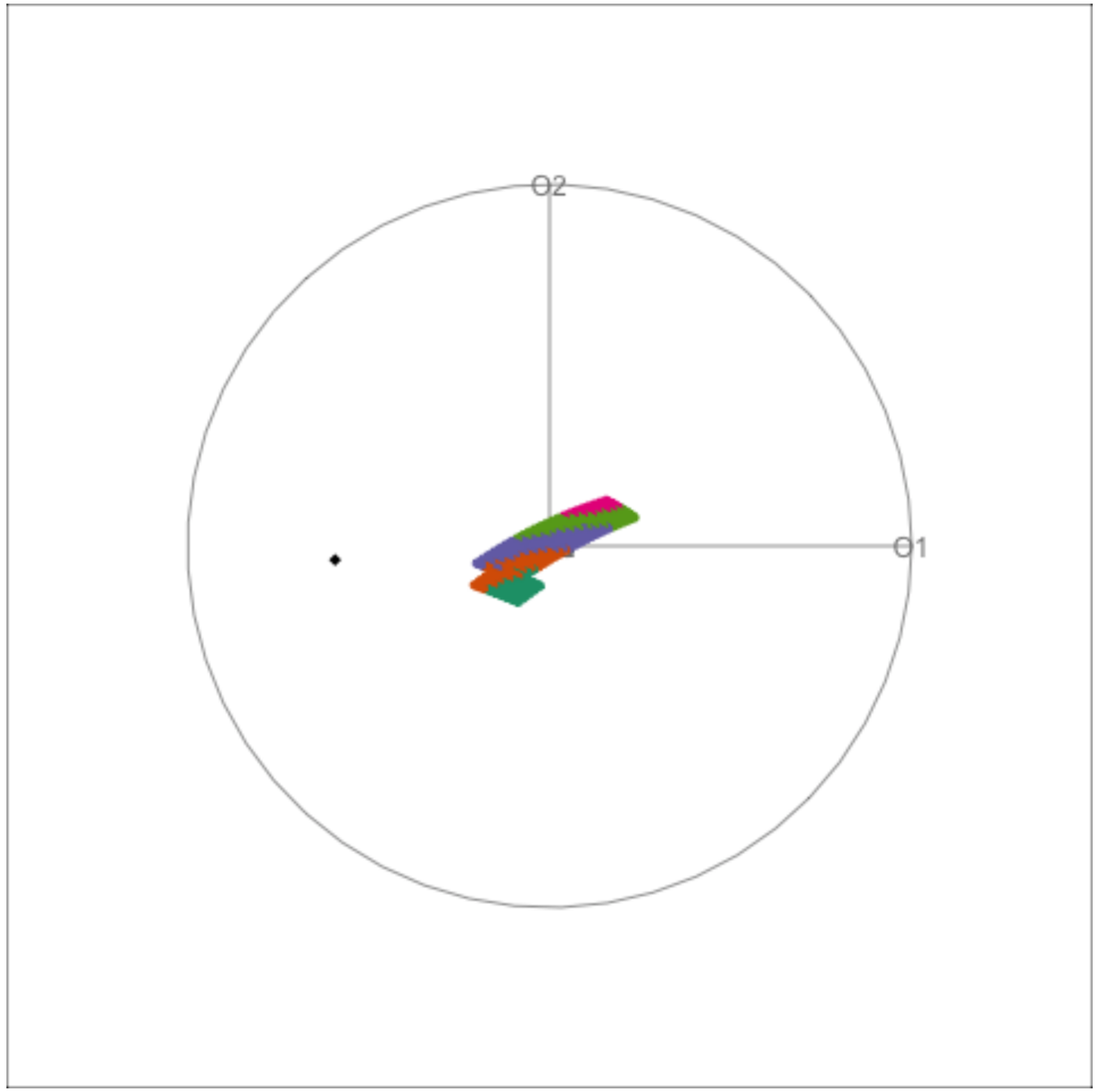
usually we see something like this



would like to see something like this

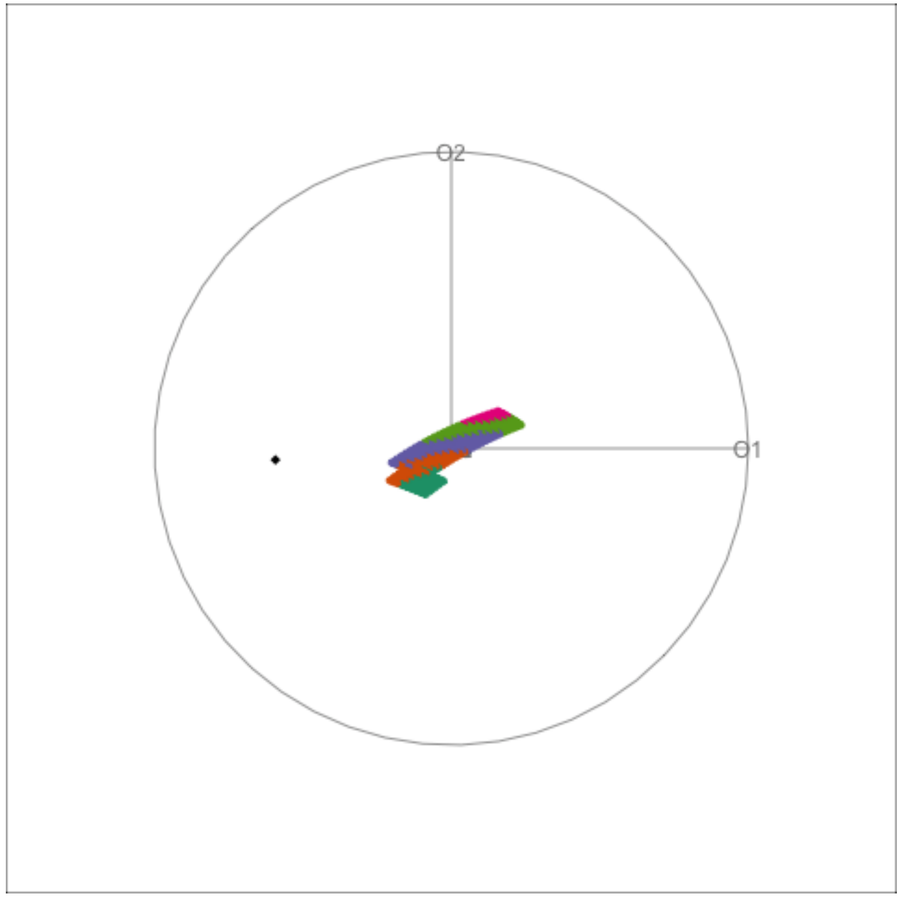


14 dimensions



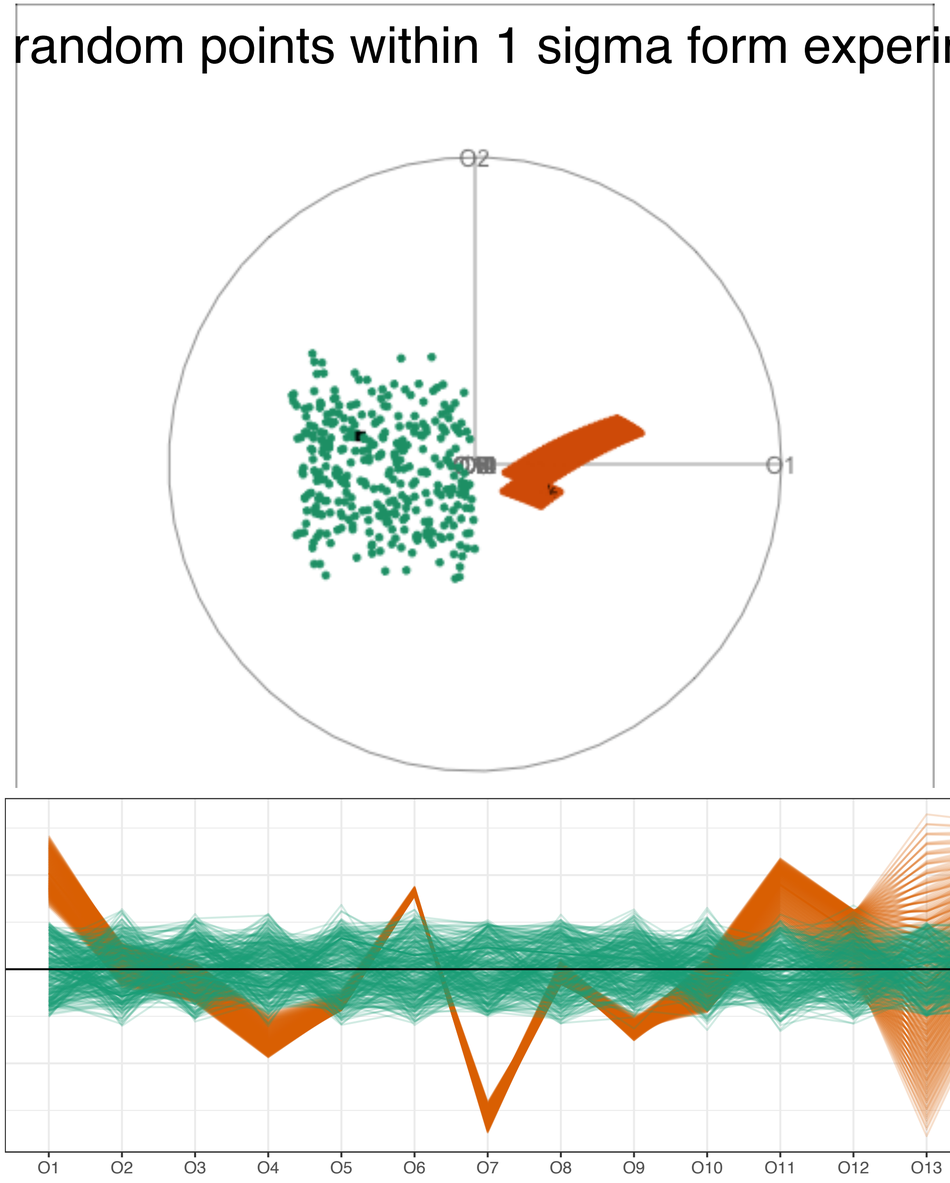
animations available at https://uschilaa.github.io/animations/ (some), or running the app interactively

grand tour display of observable space



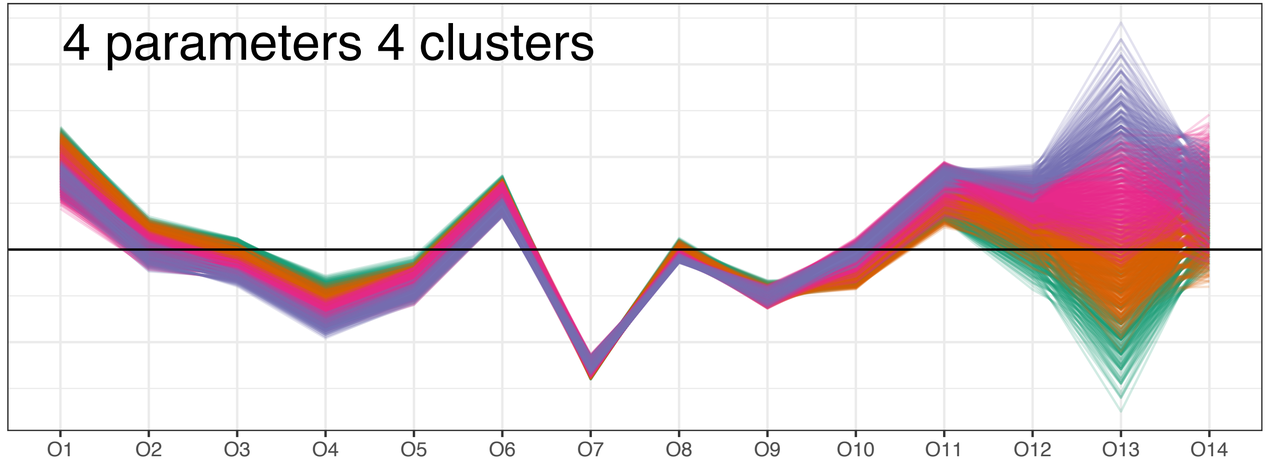
black point is experiment

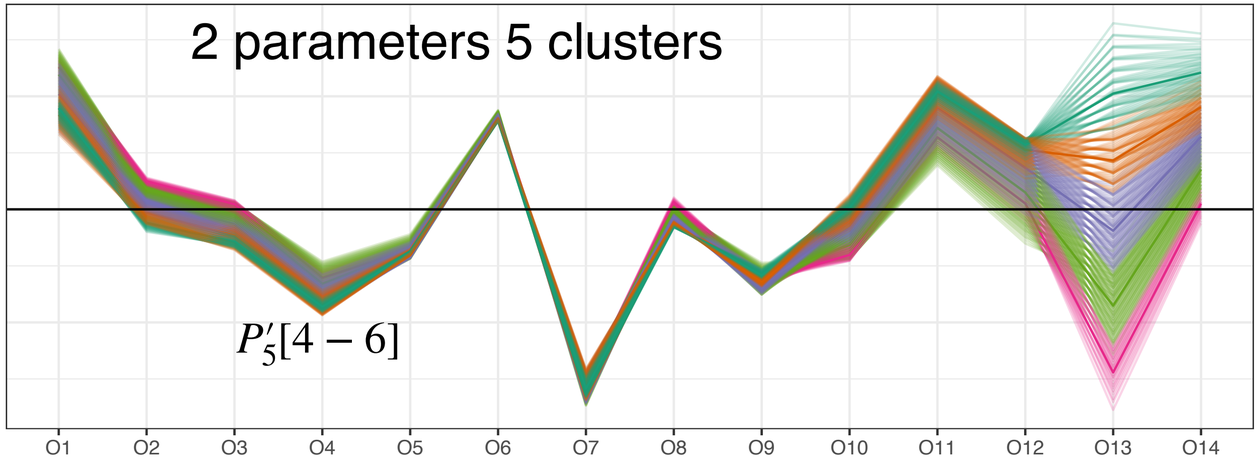
green points: random points within 1 sigma form experiment



increase the number of parameters to 4

include $C_{9''}^{\mu}$, $C_{10'}^{\mu}$

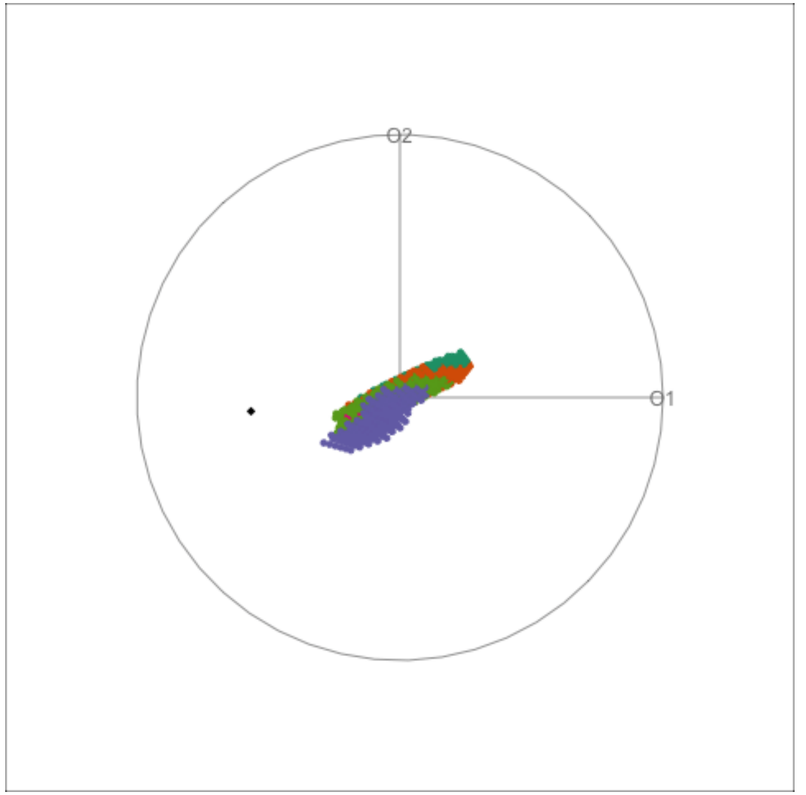




- resolution is 4 clusters
  - now 1σ cluster radius/separation for four degrees of freedom
  - Clusters now depend on 4d volume
  - PC plot showing clusters with Euclidean distance and ward.D2 linkage
- extended range of predictions with the two additional parameters increases overlap with experiment
  - 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 space visualisation

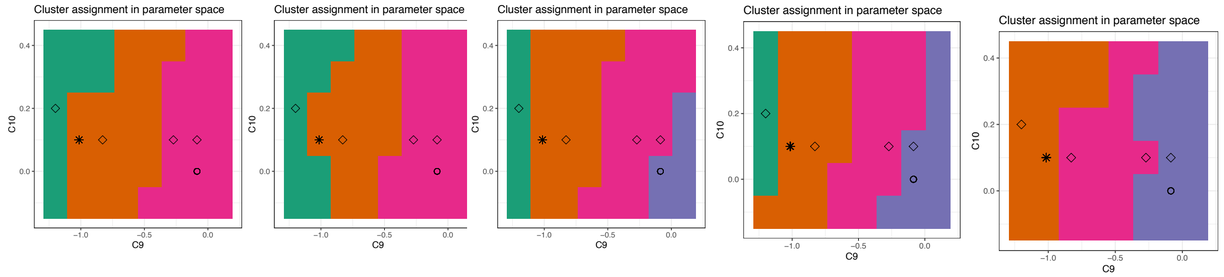
- to view the partitioning of parameter space with more than two parameters requires visualising a second high-dimensional space
- preferred solution: slice tour, but not implemented yet (see Ursula Laa, Dianne Cook, G. V., Journal of Computational and Graphical Statistics 29 (2020) 681-687 • e-Print: 1910.10854 and e-Print: 2004.13327)
- if the models are scanned on a grid we can produce a series of slices to get an idea
- the slices may reveal correlations in subspaces as illustrated in the next slide



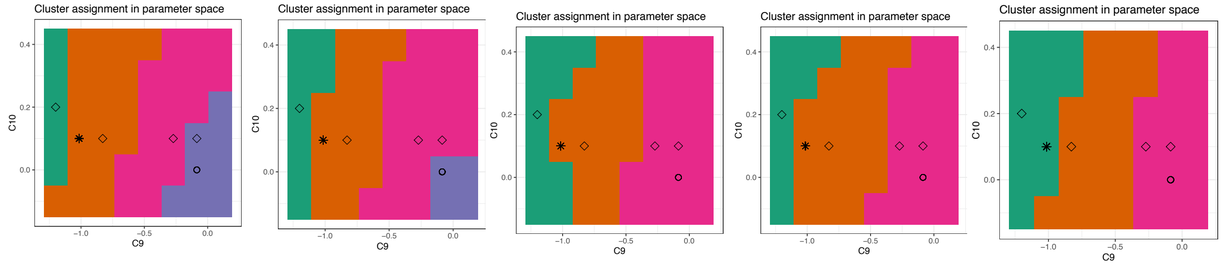
tour in observable space

slices showing the C_9^μ, C_{10}^μ plane

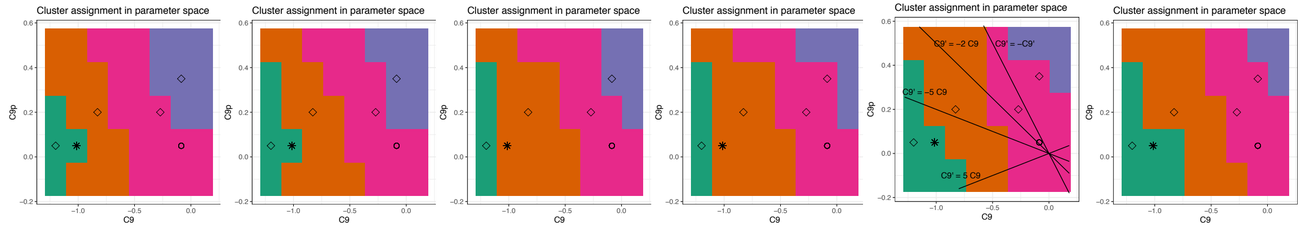
- partitioning shows some but mild dependence on $C_{9'}^\mu$, $C_{10''}^\mu$ as expected



$$C_{9'} = -0.1,\ 0.05,\ 0.2,\ 0.35,\ 0.5, \quad C_{10'} = 0$$



$$C_{10'} = -0.2,\ -0.1,\ 0,\ 0.1,\ 0.2, \quad C_{9'} = 0.05$$



slices showing the C_9^μ, $C_{9'}^\mu$ plane

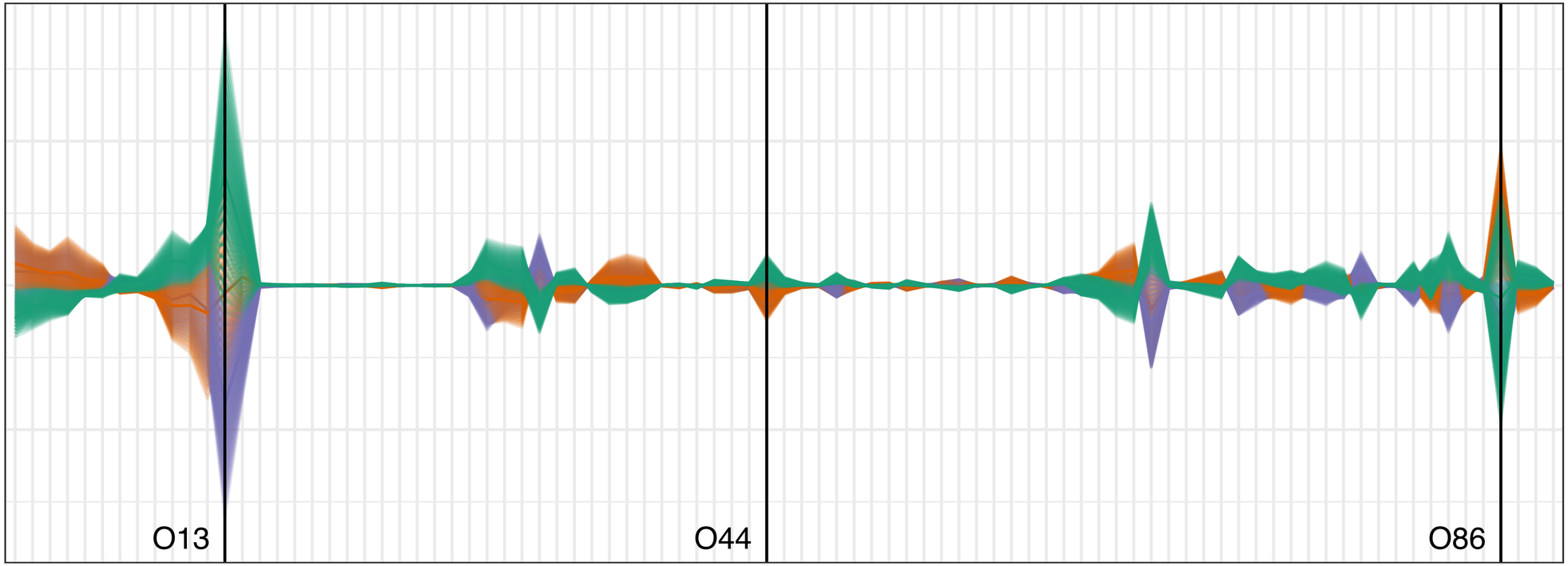
$$C_{10} = -0.1,\ 0,\ 0.1,\ 0.2,\ 0.3,\ 0.4, \quad C_{10'} = 0$$

- exhibits correlations in this plane

Include more observables

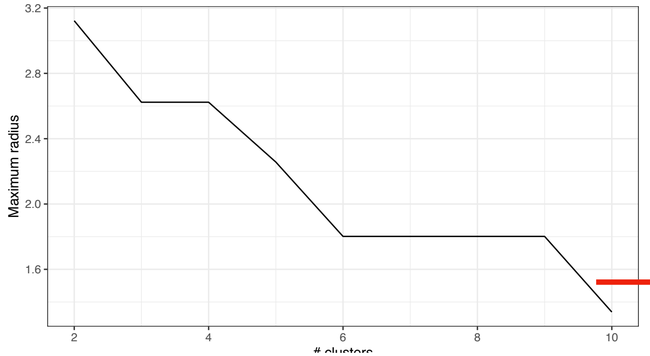
larger observable set

- 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

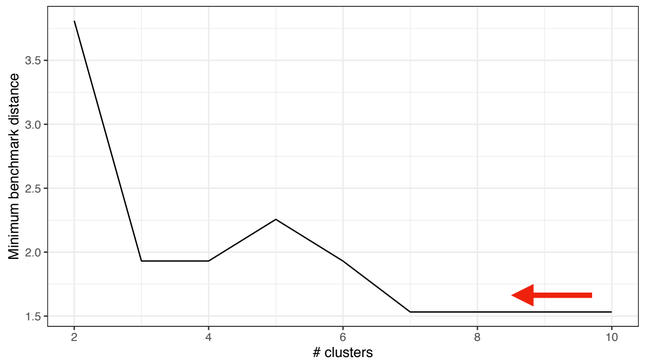


additional resolving power

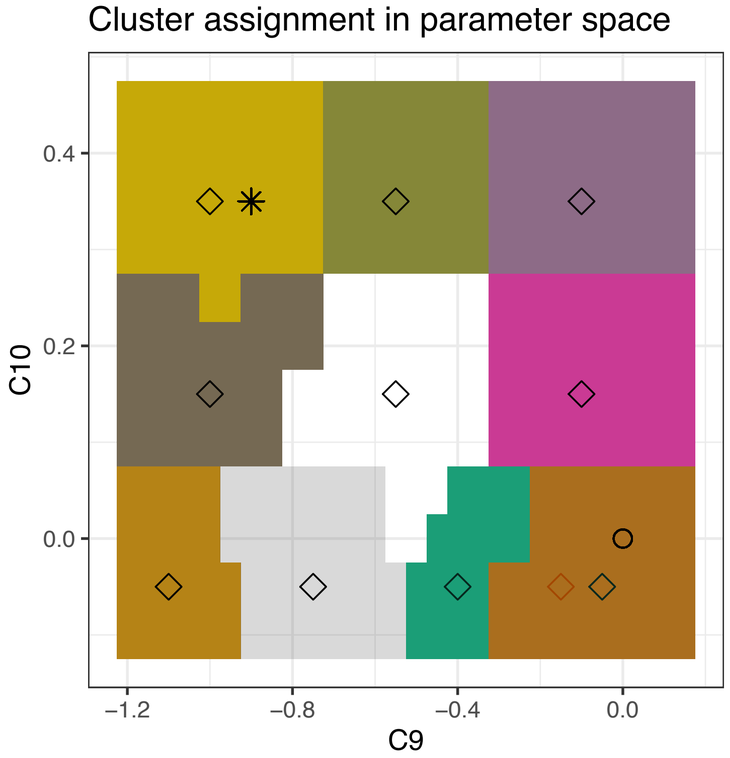
- with 89 observables the resolving power increases to about 8-10 clusters
- combined resolution in both C_9^μ and C_{10}^μ
- numerical precision is important here, these results are only indicative



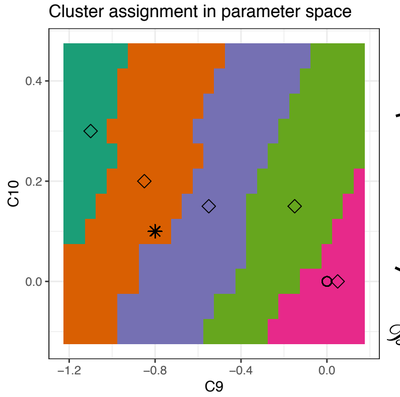
$$r \lesssim 1.5 \implies n \gtrsim 9$$



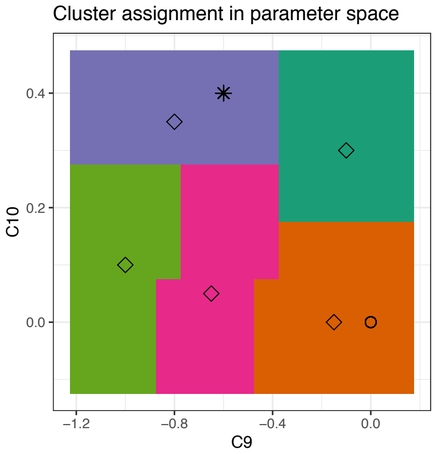
$$d \gtrsim 1.5 \implies n \lesssim 10$$

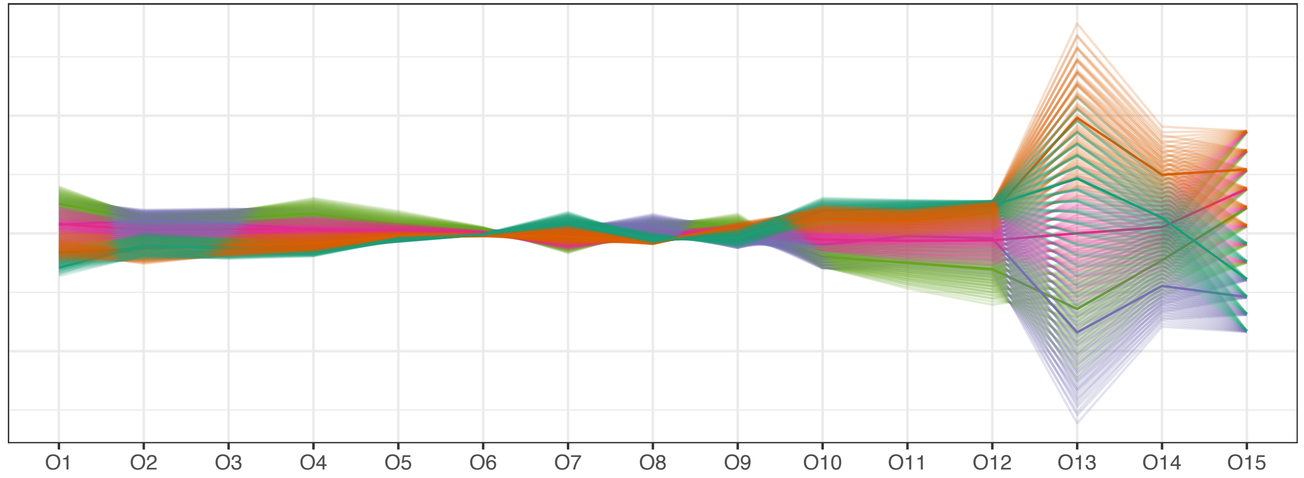


effect of (86) $\mathscr{B}(B_s \to \mu^+\mu^-)$ on original set



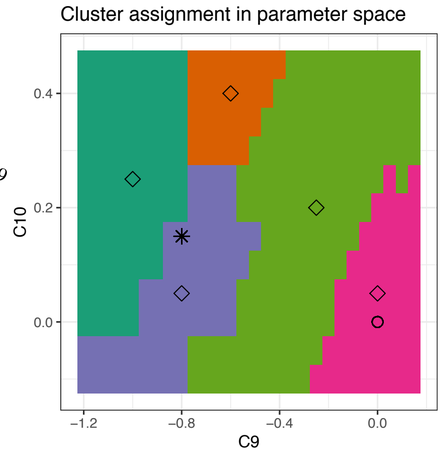
$\mathscr{B}(B_s \to \mu^+\mu^-) = (2.81^{+0.24}_{-0.22}) \times 10^{-9}$

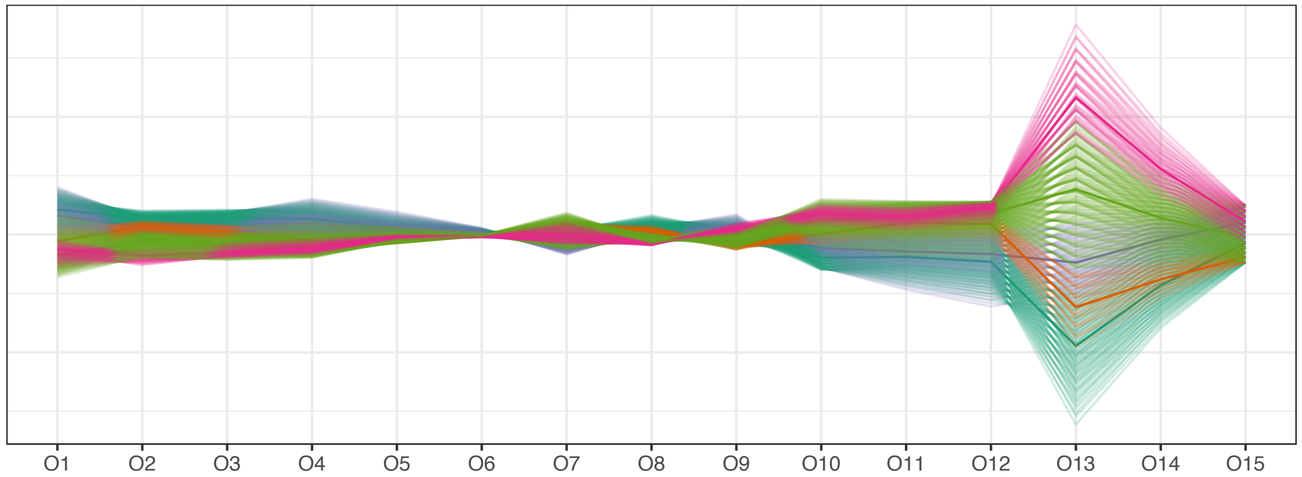


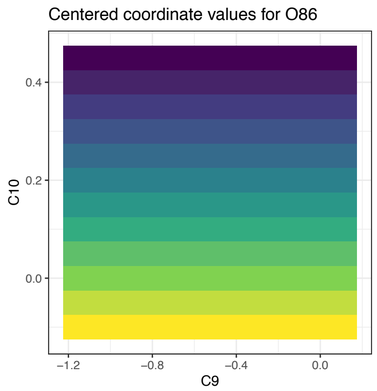


$O_{15} = \mathscr{B}(B_s \to \mu^+\mu^-)$

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

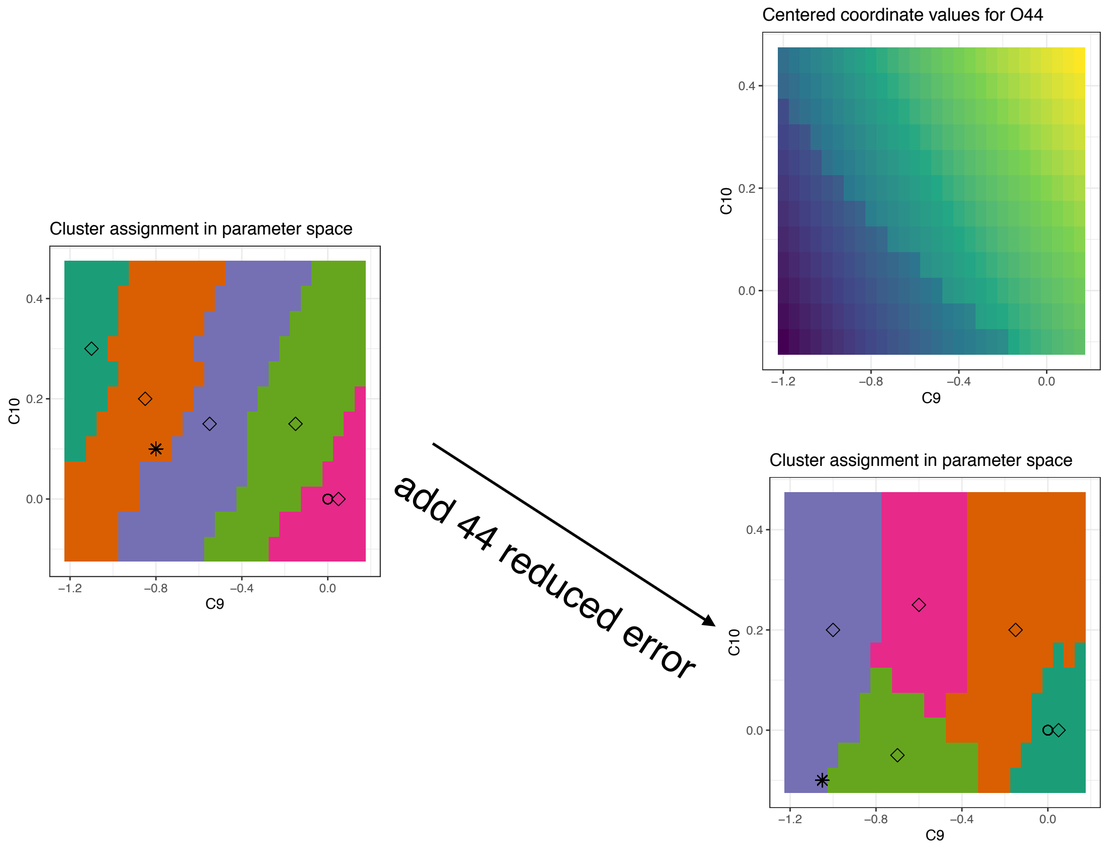


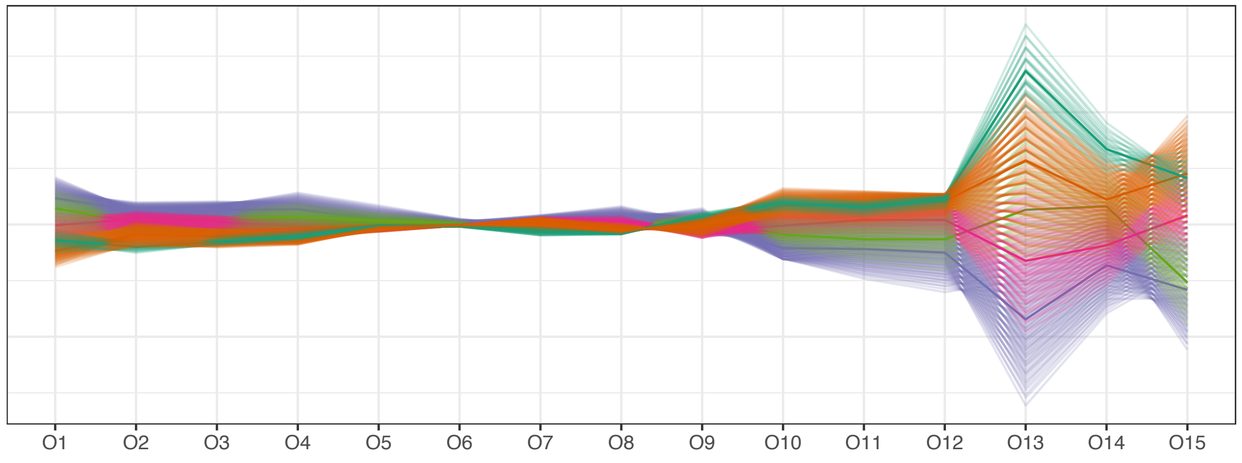




effect of (44) $P_4'[0.1 - 0.98]$ on original set

- 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





impact of future measurements

Table 67. The Belle II sensitivities to $B \to K^{(*)}\ell^+\ell^-$ observables that allow testinf of lepton flavor universality. Some numbers at Belle are extrapolated to 0.71 ab^{-1}.

Observables	Belle 0.71 ab^{-1}	Belle II 5 ab^{-1}	Belle II 50 ab^{-1}
Q_2 ([1.0, 2.5] GeV^2)	0.64	0.24	0.080
Q_2 ([2.5, 4.0] GeV^2)	0.60	0.22	0.072
Q_2 ([4.0, 6.0] GeV^2)	0.48	0.18	0.058
Q_2 (>14.2 GeV^2)	0.17	0.068	0.022
Q_3 ([1.0, 2.5] GeV^2)	0.64	0.24	0.080
Q_3 ([2.5, 4.0] GeV^2)	0.60	0.22	0.072
Q_3 ([4.0, 6.0] GeV^2)	0.48	0.18	0.058
Q_3 (>14.2 GeV^2)	0.36	0.14	0.044
Q_4 ([1.0, 2.5] GeV^2)	1.0	0.36	0.11
Q_4 ([2.5, 4.0] GeV^2)	0.90	0.30	0.10
Q_4 ([4.0, 6.0] GeV^2)	0.68	0.24	0.080
Q_4 (>14.2 GeV^2)	0.52	0.20	0.064
Q_5 ([1.0, 2.5] GeV^2)	0.94	0.34	0.11
Q_5 ([2.5, 4.0] GeV^2)	0.84	0.30	0.10
Q_5 ([4.0, 6.0] GeV^2)	0.68	0.24	0.080

PTEP

Prog. Theor. Exp. Phys. **2019**, 123C01 (654 pages)
DOI: 10.1093/ptep/ptz106

The Belle II Physics Book

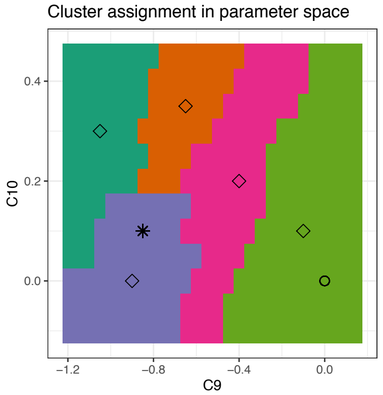
- study effect of a small set of proposed new observables (Capdevila et. al JHEP 10 (2016) 075) for which:
  - there exist sensitivity projections for Belle II
  - were singled out as important by the pull and residual analysis *Eur.Phys.J.C* 79 (2019) 6, 462

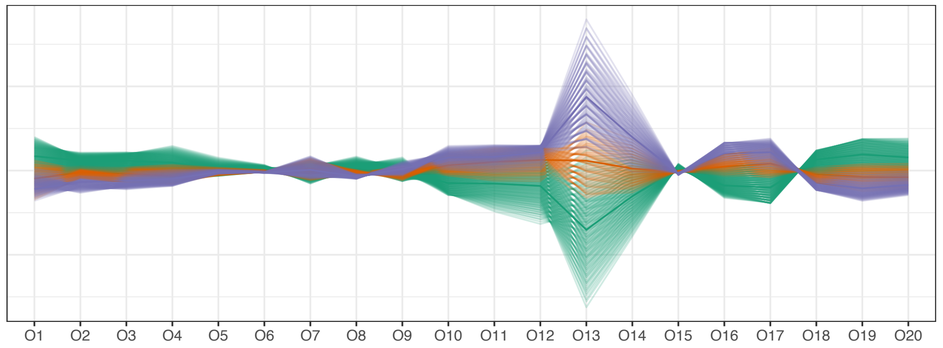
$$Q_2 \equiv P_2^\mu - P_2^e, \qquad Q_5 \equiv P_5'^\mu - P_5'^e.$$

ID	Observable	ID in [4]	SM prediction from `flavio`
1	$Q_2(B \to K^*)[1.1, 2.5]$	20	$-(6.83 \pm 0.32) \times 10^{-3}$
2	$Q_2(B \to K^*)[2.5, 4]$	26	$-(0.16 \pm 0.71) \times 10^{-3}$
3	$Q_2(B \to K^*)[4, 6]$	32	$(1.37 \pm 0.36) \times 10^{-3}$
4	$Q_5(B \to K^*)[1.1, 2.5]$	22	$(0.72 \pm 1.3) \times 10^{-3}$
5	$Q_5(B \to K^*)[2.5, 4]$	28	$-(3.72 \pm 0.67) \times 10^{-3}$
6	$Q_5(B \to K^*)[4, 6]$	34	$-(3.45 \pm 0.30) \times 10^{-3}$

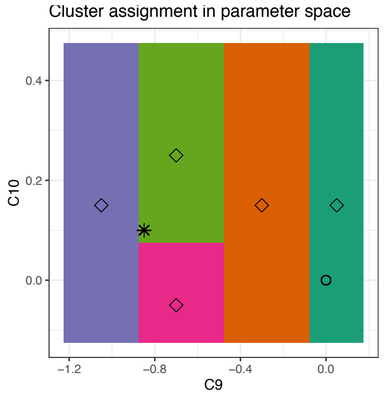
combined 20 observables

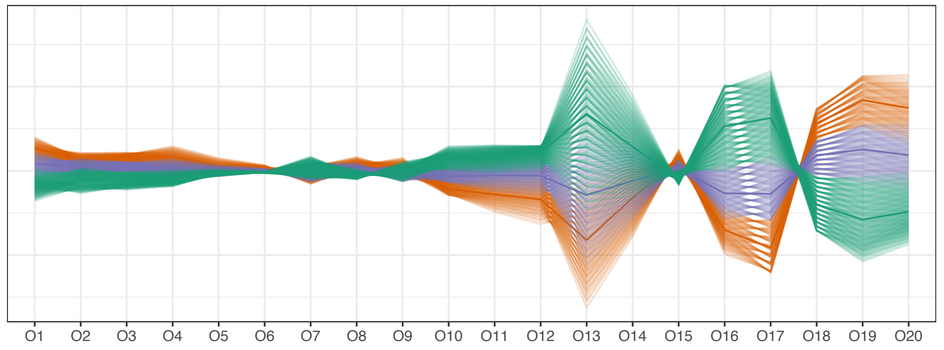
- assume future measurements will fall on current best fit (affects best fit but not partitioning)
- slightly higher resolution, 6 clusters
- can project the effect of different scenarios on future global analysis





With 5 ab^{-1} R_K still dominates

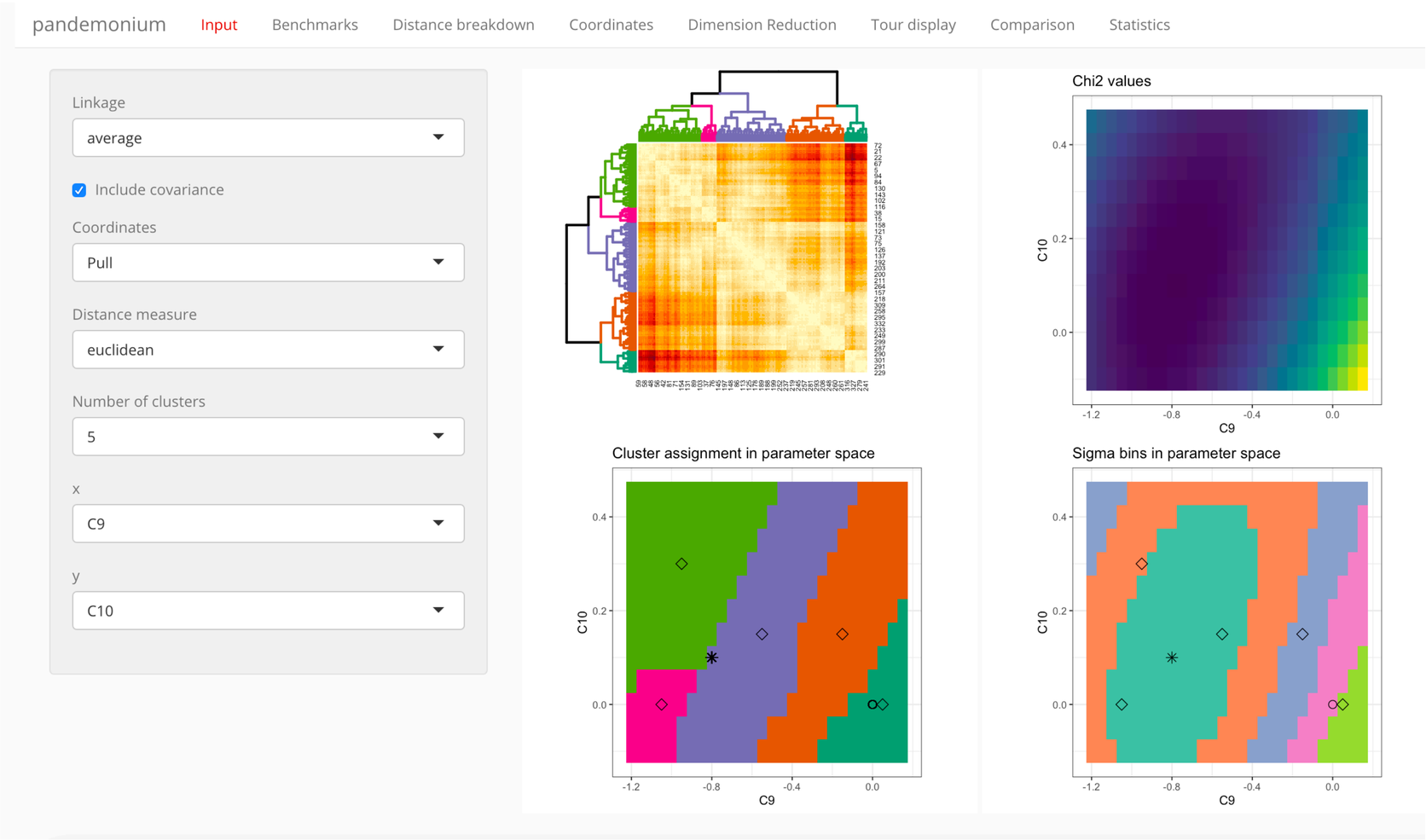




With 50 ab^{-1}

implementation by **Ursula Laa** available on GitHub in the form of an R package

https://github. com/uschiLaa/pandemonium



Thanks