Web-based Supplementary Materials for "Circular Piecewise Regression with an Application to Cell-cycle Biology" by C. Rueda, M. A. Fernández, S. Barragán, Kanti V. Mardia and Shyamal D. Peddada

1   Web Appendix A

In this section, we detail the algorithm for the computation of the estimated generalized degrees of freedom (GDF).

Algorithm for GDF estimation

- Repeat $t = 1, \ldots, T$.

- Generate N independent random values δ_{tij}, with $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$, from a $M(0, \tau)$ distribution and denote $\Delta_t = (\delta_{tij})$.

- Evaluate the fitted values $\hat{\mu}_{ij}(\boldsymbol{\Psi} + \Delta_t)$ using the perturbated data $\boldsymbol{\Psi} + \Delta_t$ under the piecewise model.

- For each pair (i, j), calculate $\hat{h}_{ij}$ as the regression slope from

$$\hat{\mu}_{ij}(\boldsymbol{\Psi} + \Delta_t) = \alpha_{ij} + \hat{h}_{ij}\delta_{tij}, \ t = 1, \ldots, T.$$

- Estimate $GDF = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{\partial \hat{\mu}_{ij}}{\partial \psi_{ij}}$ as $\widehat{GDF} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \hat{h}_{ij}$.

As recommended by Ye [1998], in our implementation of this algorithm we have considered values for the tuning parameter τ such that κ/τ is bounded away from 0 and a number of perturbations $T \geq N$.

2   Web Appendix B

In this section, we present the data (Web Table 1) for the example developed in Section 3.1 (Within species between labs correlation of phase angles of cell-cycle genes) of the main paper. Also some figures related to that example are given. To be more precise Web Figure 1 shows a scatterplot of the data while the residuals of the piecewise and circular-circular fits from the Downs and Mardia [2002] model appear in Web Figure 2.

Notice that the residual plots indicate that two of the points with high residual are better adjusted by the piecewise model while a third one is not, otherwise there are no major changes in the residuals.

Gene	Sector 1 θ	ψ	Gene	Sector 3 θ	ψ
MSH6	1.8318	2.3902	KIP3	4.2105	4.2793
PST1	2.0674	0.8357	FKH1	4.4881	4.0768
SMC3	2.1899	2.6578	MYO1	4.668	4.8854
RFA1	2.3153	2.5433	SWI5	5.0686	4.3735
MRC1	2.321	2.8085	HOF1	5.0963	4.908
POL2	2.4471	2.7235	MOB1	5.1074	5.1097
MCD1	2.5083	2.5613	SST2	5.1544	2.8915
SWE1	2.6909	2.8864	BUD4	5.254	4.6404
CLN2	2.7518	2.6634	CHS2	5.3986	4.7619
DSE4	2.7657	2.6942	ASE1	5.4247	5.3721
			CDC5	5.5488	4.6247
			KIN3	5.6826	5.3336

| Gene	Sector 2 θ	ψ	Gene	Sector 4 θ	ψ
RAD51	2.9231	2.4926	CDC20	6.0639	5.5172
HHT1	3.0139	3.6616	DBF2	0.3716	0.5854
HHT2	3.2281	3.6451	POL1	0.5398	2.7921
HHF1	3.5405	3.5068	CDC6	0.8392	1.8843
HTB2	3.8346	3.2778			
HTA2	3.8744	3.7854			

Web Table 1: Values in each sector for the independent (θ) and the dependent (ψ) variables for the *S. Cerevisiae* data

3   Web Appendix C

In this section, we present the data (Web Table 2) for the example developed in Section 3.2 (Between species and between labs correlation of phase angles of cell-cycle genes) of the main paper. Notice that gene $BUD4$ considered in the other example was not considered in this case because it was not very periodic in *S. pombe*. In the case of gene $HOF1$ for *S. pombe* we used two paralogs (namely *cdc15* and imp2) since they are both highly periodic.

Also some figures related to that example are given here. Web Figure 3 shows a scatterplot of the data while the residuals of the piecewise and circular-circular fits appear in Web Figure 4.

In this case, a careful analysis of the residuals suggests that there may be 3 potential outliers in the
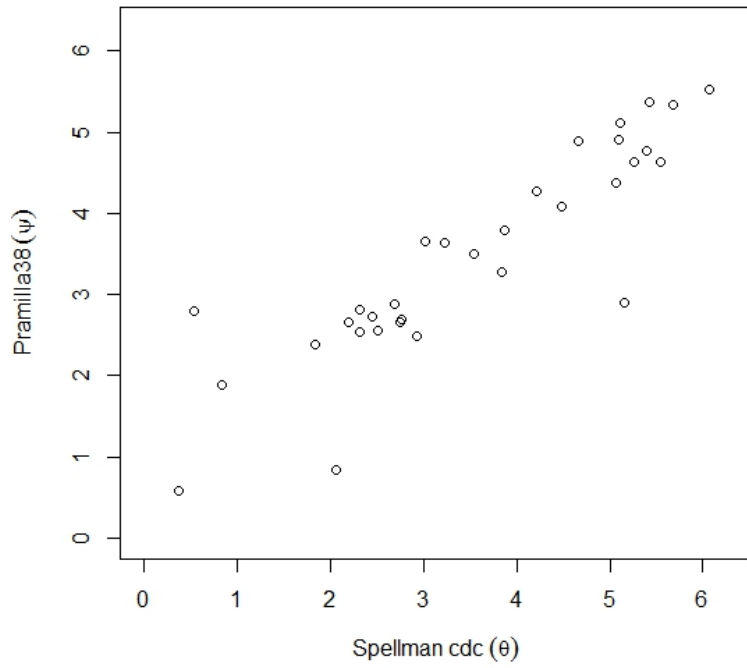
Web Figure 1: Scatterplot of the Spellman cdc vs the Pramilla38 data
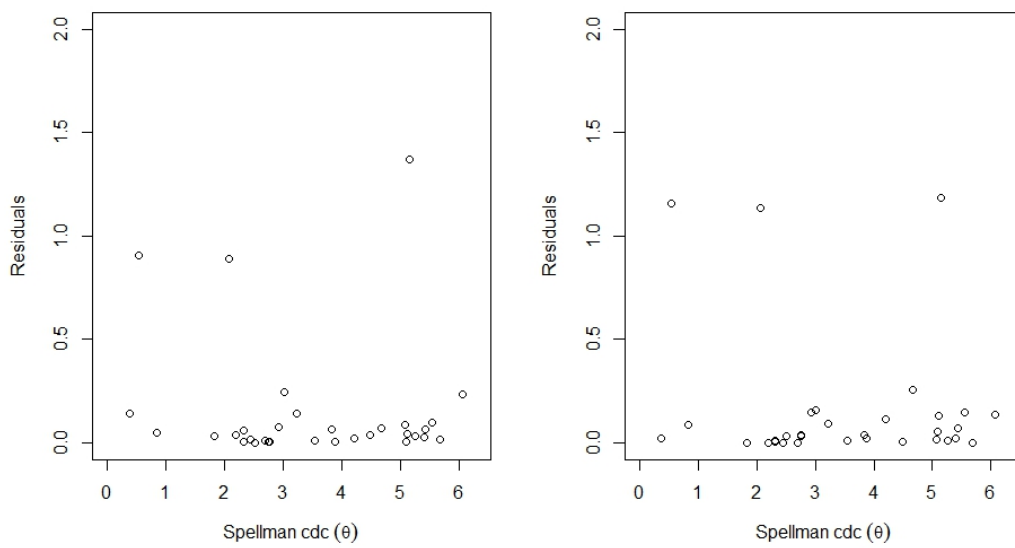
Web Figure 2: Residuals vs θ for the *S. Cerevisiae* data (piecewise at left and c-c at right)

data. These three points with high residual are only slightly better adjusted by the piecewise model. The improvement obtained by the piecewise model is more apparent for the non outliers as, for example, there are no more points with residual higher than 1 for the piecewise analysis while there are two such points in the c-c fit. Although further biological evaluations and confirmations are necessary, the three potential outliers identified by our piecewise regression, namely, MCD1, CHS2 and ASE1, may have different phases of peak expression between the two species of yeasts. Thus, in this example, our method was sensitive to identify features in the data that could potentially stimulate biologists to explore new hypotheses to test.

Cerevisiae Gene	Sector 1 θ	ψ	Cerevisiae Gene	Sector 3 θ	ψ
SMC3	2.1899	6.0886	KIP3	4.2105	4.9841
RFA1	2.3153	5.9642	FKH1	4.4881	3.6002
MRC1	2.321	5.6791	MYO1	4.668	3.297
POL2	2.4471	0.5917	SWI5	5.0686	4.858
MCD1	2.5083	3.8476	HOF1	5.0963	5.2881
SWE1	2.6909	6.199	HOF1	5.0963	5.2981
CLN2	2.7518	5.6399	MOB1	5.1074	6.1698
DSE4	2.7657	6.0967	SST2	5.1544	5.5035
			CHS2	5.3986	2.6417
			ASE1	5.4247	2.4665
			CDC5	5.5488	5.0901
			KIN3	5.6826	3.3911

| Cerevisiae Gene	Sector 2 θ	ψ	Cerevisiae Gene	Sector 4 θ	ψ
RAD51	2.9231	1.3642	CDC20	6.0639	5.6069
HHT1	3.0139	0.9978	DBF2	0.3716	4.8342
HHT2	3.2281	0.6728	POL1	0.5398	6.0462
HHF1	3.5405	1.0577	CDC6	0.8392	5.3213
HTB2	3.8346	1.1463	MSH6	1.8318	5.4055
HTA2	3.8744	1.0249	PST1	2.0674	5.3788

Web Table 2: Values in each sector for the independent (θ) and the dependent (ψ) variables for the two species data

References

T.D. Downs and K.V. Mardia. Circular regression. *Biometrika*, 89(3):683–697, 2002.

J. Ye. On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, 93(441):120 – 131, 1998.
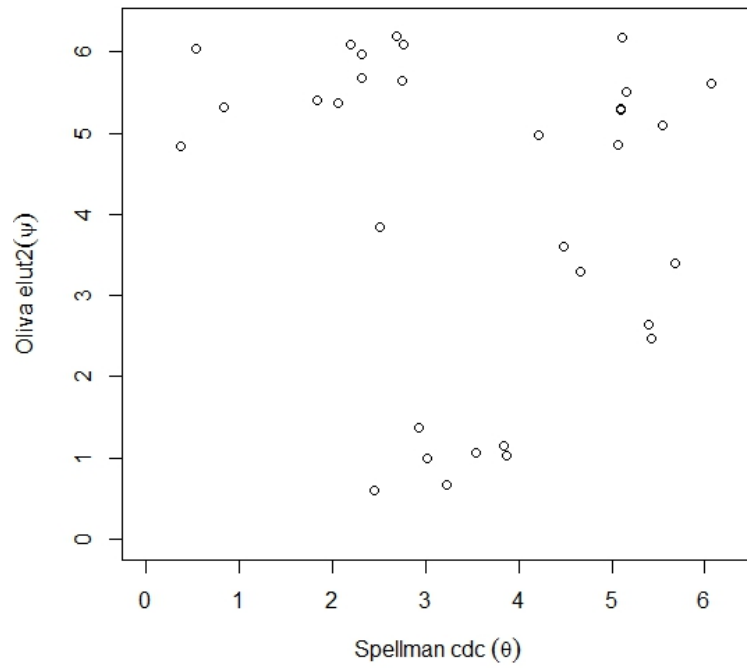
Web Figure 3: Scatterplot of the Spellman cdc (*S. cerevisiae*) vs the Oliva elut2 (*S. pombe*) data
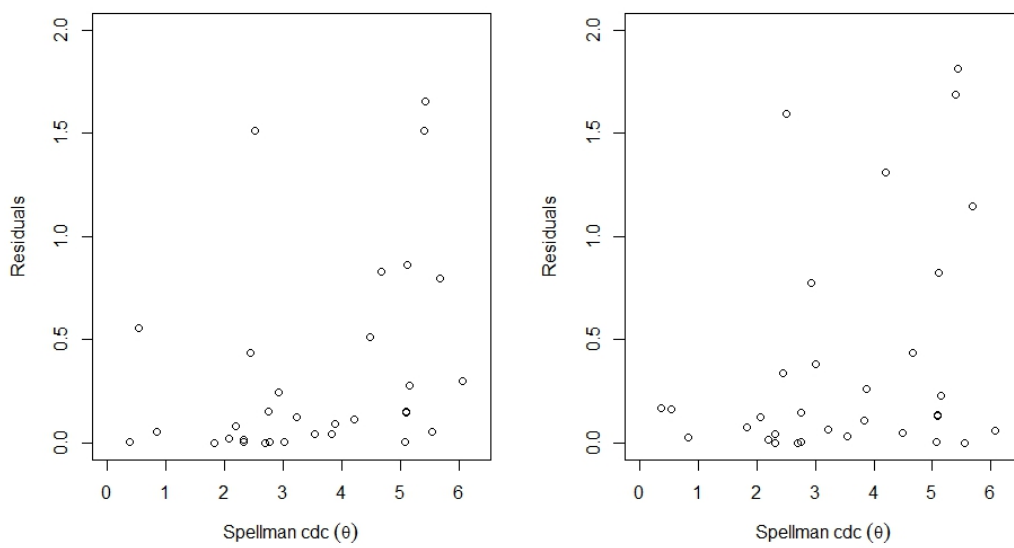
Web Figure 4: Residuals vs θ for the two species data (piecewise at left and c-c at right)