

1 **Estimations of the weather effects on brain functions using functional MRI – a cautionary tale**

2 Xin Di ^{1,2}, Marie Wolfer ^{1,3,4}, Simone Kühn ^{5,6}, Zhiguo Zhang ^{7,8}, Bharat B. Biswal ^{1,2,*}

3

4 1. Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ, 07102, USA

5 2. School of Life Sciences and Technology, University of Electronic Science and Technology of China,

6 Chengdu, China

7 3. Clinical Affective Neuroimaging Laboratory (CANLAB), Otto-von-Guericke-University Magdeburg,

8 Magdeburg, Germany

9 4. Department for Behavioral Neurology, Leibniz Institute for Neurobiology, Magdeburg, Germany

10 5. Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

11 6. Clinic and Polyclinic for Psychiatry and Psychotherapy, University Clinic Hamburg-Eppendorf,

12 Germany

13 7. School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China

14 8. Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging,

15 Shenzhen, China

16

17 * Corresponding author:

18 Bharat B. Biswal, PhD

19 607 Fenster Hall, University Height

20 Newark, NJ, 07102, USA

21 bbiswal@yahoo.com

22

23 **Abstract**

24 The influences of environmental factors such as weather on human brain are still largely unknown. A few
25 neuroimaging studies have demonstrated seasonal effects, but were limited by their cross-sectional design
26 or sample sizes. Most importantly, the stability of MRI scanner hasn't been taken into account, which
27 may also be affected by environments. In the current study, we analyzed longitudinal resting-state
28 functional MRI (fMRI) data from eight individuals, where the participants were scanned over months to
29 years. We applied machine learning regression to use different resting-state parameters, including
30 amplitude of low-frequency fluctuations (ALFF), regional homogeneity (ReHo), and functional
31 connectivity matrix, to predict different weather and environmental parameters. For a careful control, the
32 raw EPI and the anatomical images were also used in the prediction analysis. We first found that daylight
33 length and temperatures could be reliability predicted using cross-validation using resting-state
34 parameters. However, similar prediction accuracies could also achieved by using one frame of EPI image,
35 and even higher accuracies could be achieved by using segmented or even the raw anatomical images.
36 Finally, we verified that the signals outside of the brain in the anatomical images and signals in phantom
37 scans could also achieve higher prediction accuracies, suggesting that the predictability may be due to the
38 baseline signals of the MRI scanner. After all, we did not identify detectable influences of weather on
39 brain functions other than the influences on the stability of MRI scanners. The results highlight the
40 difficulty of studying long term effects on brain using MRI.

41

42 **Keywords:** daylight length, environmental effects on brain, machine learning regression, resting-state,
43 scanner stability, temperature, weather.

44

45 **1. Introduction**

46 Daily environmental factors such as weather and seasonality affect mood and cognitive functions (Keller
47 et al. 2005; Denissen et al. 2008; Cedeño Laurent et al. 2018; IJzerman et al. 2018; Lim et al. 2018), and
48 may lead to pathological affective disorder (Kurlansik and Ibay 2012). The effects on individuals may be
49 small, but the collective effects may lead to broader impacts, e.g. on stock markets (Saunders 1993;
50 Hirshleifer and Shumway 2003). To better understand the effects of weather and seasonality on mood or
51 cognitions, it is critical to study their effects on brain functions. A few human neuroimaging studies have
52 explored this association. Seasonal effects on brain functions as measured by functional MRI (fMRI)
53 have been observed both in resting-state (Choe et al. 2015) and when performing cognitive tasks (Meyer
54 et al. 2016). Some neural transmitter activity in the striatum also showed seasonal effects, i.e. serotonin
55 transmitter binding as measured by ^{11}C -labeled 3-amino-4-(2-dimethylaminomethyl-phenylsulfanyl)-
56 benzonitrile (^{11}C]DASB) positron emission tomography (PET) (Praschak-Rieder et al. 2008; Kalbitzer et
57 al. 2010; Mc Mahon et al. 2016) and dopamine synthesis as measured by ^{18}F -DOPA PET (Eisenberg et al.
58 2010; Kaasinen et al. 2012). A study even reported seasonal changes of hippocampal volumes in human
59 subjects (Miller et al. 2015).

60 There are several limitations in these neuroimaging studies. First, most of these studies are cross-
61 sectional, which is limited by the large individual differences in brain functions (Gordon et al. 2017). In
62 addition, most of the studies examined roughly defined seasonal effects or yearly periodical effects. But
63 the exact phase of the seasonal variations may be different from the roughly defined four seasons.
64 Sometimes the yearly effects showed different phases (Meyer et al. 2016), suggesting more complicated
65 relationships of environmental factors on brain functions. Therefore, it is critical to examine which
66 environmental parameters such as weather have more contributions to the seasonal effects. Among
67 different weather parameters, daylight length and temperature represent the significant environmental
68 differences in seasonal fluctuations. Gillihan et al has explored the weather effects on brain functions
69 using a small cross-sectional sample (Gillihan et al. 2011). They identified a weather index related to

70 mood, and showed that the weather index was correlated with resting-state cerebral blood flow as
71 measured by arterial spin labeled (ASL) perfusion fMRI mainly in the insula. But more systematic
72 examinations of weather effects on brain functions have not been performed. Lastly, the most commonly
73 used neuroimaging method is blood-oxygen-level dependent (BOLD) signal based fMRI, where the
74 interpretation of the results should consider neuronal level, neurophysiological level, and the underlying
75 physical level of the scanner. Specifically, if some effects on fMRI signals were observed, they may be
76 due to the changes of neuronal activity, which is favorable to psychologists and psychiatrist. But the
77 effects may also due to the changes of neurovascular coupling (Di et al. 2013; Yuan et al. 2013), changes
78 of brain structures, or even the stability of the MRI scanner. Therefore, when examining the weather
79 effects on brain functions, the alternative factors need to be considered and carefully controlled.

80 The purpose of the current study is to estimate to what extent resting-state brain functions were
81 affected by weather. We analyzed longitudinal resting-state fMRI data from eight individuals from three
82 datasets, where the individuals were scanned over periods of months to years (Choe et al. 2015; Poldrack
83 et al. 2015; Filevich et al. 2017). One challenge for estimating weather effects is that the effects may be
84 small. Therefore, we applied a machine learning regression approach to evaluate the amount of effects.
85 Because multiple brain regions have been implicated in seasonal effects, e.g. basal ganglia (Praschak-
86 Rieder et al. 2008; Kalbitzer et al. 2010; Mc Mahon et al. 2016), insula (Gillihan et al. 2011), and
87 hippocampus (Miller et al. 2015), small regional effects may be aggregated into detectable effects using
88 machine learning technique. We performed within-subject prediction analysis at single subject level. We
89 asked what weather parameters have the most effects on resting-state brain functions, which can be
90 represented as high prediction accuracies in predictions of these parameters. In order to rule out possible
91 confounding effects that might give rise to prediction, we also performed control prediction analyses.
92 First, we analyzed anatomical MRI images to check whether the observed prediction could be attributed
93 to anatomical variations. Second, we checked images from phantom data to examine whether the
94 prediction could be attributed to the stability of the MRI scanner.

95

96 2. Materials and methods:

97 2.1. MRI datasets

98 Several multi-session resting-state fMRI datasets were pooled together, where the subjects were scanned
99 over periods of months to years. The first subject was derived from the Kirby sample (Choe et al. 2015),
100 where the single subject was scanned for 156 sessions over three and half years. The second subject was
101 from the Myconnectome sample (Poldrack et al. 2015), where the subject was scanned 90 times over one
102 and half years. The remaining six subjects were from the Day2day sample (Filevich et al. 2017), where
103 the subjects were scanned over a similar span of about half a year. The detailed subject and scan
104 information is listed in Table 1.

	Dataset	Sex	Age	# of sessions	First scan	Last scan	# of volumes	TR (s)	Voxel size (mm ³)
1	Kirby	M	40	156	2009/12/07	2013/06/20	198	2	3 x 3 x 4
2	MyConnectome	M	45	83	2012/10/23	2014/03/11	500	1.16	2.4 x 2.4 x 2.4
3	Day2day	F	23	50	2013/07/03	2013/12/18	148	2	3 x 3 x 3.6
4	Day2day	F	31	48	2013/07/03	2014/01/08	148	2	3 x 3 x 3.6
5	Day2day	F	29	45	2013/07/03	2014/01/27	148	2	3 x 3 x 3.6
6	Day2day	F	24	46	2013/07/02	2013/12/19	148	2	3 x 3 x 3.6
7	Day2day	M	30	39	2013/07/09	2014/02/12	148	2	3 x 3 x 3.6
8	Day2day	F	29	48	2013/07/03	2014/02/20	148	2	3 x 3 x 3.6

105

106 **Table 1** Subject and scan information for the eight subjects with longitudinal scans. The number of
107 sessions represents the effective number of sessions after dropout due to missing data or large head
108 motions. The numbers of volumes represent the number of volumes used in the analysis after removing
109 the first several volumes.

110

111 The MRI data from the Kirby sample were scanned using a 3T Philips Achieva scanner. The data
112 from the Myconnectome sample were scanned using a 3T Siemens Skyra scanner using a 32-channel
113 head coil. And the data from the Day2day project were scanned using a 3T Siemens Magnetom Trio
114 scanner using a 12-channel head coil. For each subject, resting-state fMRI data with multiple sessions
115 were available. Within a subject, the resting-state fMRI were scanned using the same imaging parameters,
116 but the parameters varied between different sites. Some essential resting-state fMRI parameters are listed
117 in Table 1. For more details, we refer the readers to the original articles.

118 High resolution anatomical MRI images were available for only a few sessions in the Kirby and
119 Myconnectome datasets for each subject. A MRI image of one session was used to register all the
120 functional images to standard Montreal Neurological Institute (MNI) space. For the Day2day dataset,
121 structural MRI images were available for all the sessions. Only the structural MRI image of the last
122 session of each subject was used to aid preprocessing of the fMRI images. All the structural images of
123 the Day2day project were also used in the control prediction analysis.

124 Lastly, we obtained MRI scanner quality assurance agar phantom data from the Day2day site.
125 The images were scanned between June 2013 and February 2014 with a weekly basis (37 sessions in
126 total). The data were acquired using gradient echo (GRE) sequence with the same coil as the one used for
127 acquisition of the human data. Two images were acquired for each session. The parameters include: TR
128 = 2000 ms; TE = 30 ms; FOV = 22 cm; matrix = 64 x 64; slice number = 28; slice thickness = 4 mm (1
129 mm gap).

130 **2.2. Environmental data**

131 The fMRI data were acquired from three different cities in two continents, Baltimore USA (Kirby),
132 Austin USA (Myconnectome), and Berlin Germany (Day2day), which reflect different types of climates.
133 The latitudes of these four cities are approximately 39 °N, 30 °N, and 52 °N, respectively. The weather
134 data for the two US cities were downloaded from (US) National Centers for Environmental Information
135 website (<https://www.ncdc.noaa.gov/cdo-web/>). The Local Climatological Data from Maryland Science

136 Center Station and Austin Camp Mabry Station were used to represent the weather for the Kirby and
137 Myconnectome datasets, respectively. We used the following measures, maximum and minimum
138 temperatures, air pressure, wind speed, humidity, and precipitation. For those with missing data, we also
139 checked Daily Summaries data from the NOAA website. The weather data for the Day2day dataset were
140 collected by the German researchers. Daily sunshine hours were not used, because they were not
141 available for the other datasets.

142 We also included daylight length in the current analysis. It was already available in the NOAA
143 Local Climatological Data. For the Day2day data, we calculated the daylight length in Berlin according
144 to its geographic location through the website of the Astronomical Applications Department of the U.S.
145 Naval Observatory computes (http://aa.usno.navy.mil/data/docs/Dur_OneYear.php). For the Day2day
146 dataset, there are also three parameters that reflect local environmental variations, i.e. scanner room
147 temperature, scanner room humidity, and scanner Helium level. These three parameters were also used in
148 the prediction analysis when using Day2day data to do prediction.

149 **2.3. MRI data processing**

150 **2.3.1. Resting-state fMRI Preprocessing**

151 Data processing and statistical analysis were performed using MATLAB (R2017b). SPM12
152 (<http://www.fil.ion.ucl.ac.uk/spm/>; RRID:SCR_007037) was used for fMRI data preprocessing. The first
153 2, 18, and 2 functional images for each session were discarded for Kirby, Myconnectome, and Day2day
154 datasets, respectively, remaining 198, 500, and 148 images for each session. For each subject, all the
155 functional images were realigned to the first session. All the prediction analysis was performed in native
156 space of each subject. The anatomical images were coregistered to the mean functional image, and then
157 segmented into gray matter (GM), white matter (WM), cerebrospinal fluid (CSF), and other tissues. At
158 each voxel, Friston's 24 head motion model, the first five principle components from WM signals and the
159 first five principle components from CSF signals were regressed out, and then a band-pass filtering was

160 applied at frequency of 0.01 and 0.1 Hz. The images were not spatially smoothed, because there was no
161 voxel-wise univariate analysis involved.

162 **2.3.2. ALFF, ReHo, and connectivity matrices**

163 We calculated three resting-state parameters to represent resting-state brain functions, i.e. amplitude of
164 low-frequency fluctuation (ALFF) (Zang et al. 2007) and regional homogeneity (ReHo) (Zang et al. 2004)
165 to represent regional properties, and connectivity matrix to represent inter-regional connectivity property.
166 ALFF and ReHo were calculated using REST toolbox (Song et al. 2011). Essentially, ALFF calculated
167 the power of the time series signals between 0.01 to 0.08 Hz at every voxel, resulting in an ALFF map for
168 each session. ReHo calculated the correlations of the current voxel with the 26 neighboring voxels, which
169 also resulted in a ReHo map for each session. The ALFF and ReHo values for each session within the
170 subject's GM mask were concatenated as a vector for further analysis. The subject specific GM masks
171 were defined as GM intensity greater than 0.5 based on the segmentation results from the subject's own
172 anatomical image. Because the GM masks were defined in the native spaces and the fMRI resolution
173 varies across datasets, the number of within mask voxels also varied (from 20,780 to 55,368).

174 Correlation matrices were calculated among 164 regions of interests (Dosenbach et al. 2010; Di
175 and Biswal 2019). Spherical ROIs were first defined in MNI space with a radius of 8 mm, and
176 transformed into the native space for each subject. There were in total 13,366 connectivity values ($164 \times$
177 $(164 - 1) / 2$), which were concatenated as a vector for the prediction analysis. The correlation values
178 were transformed into Fishers' z scores.

179 **2.3.3. Head motion and other potential confounding variables**

180 To minimize the confounding of head motion in the prediction analysis, we first removed sessions with
181 large head motions. We calculated frame-wise displacement in translation and rotation directions (Di and
182 Biswal 2015). A session's data with maximum frame-wise displacement greater than 1 mm or 1° were
183 discarded. No sessions were removed in the Kirby data, and seven sessions (7.8%) were removed for the
184 Myconnectome data. In the Day2day dataset, at most two sessions were removed for each subject.

185 Secondly, we regressed out 24 motion variables using Friston's head motion model (Friston et al. 1996),
186 which has been shown to be effective to minimize the effects of head motion on resting-state measures
187 (Yan, Cheung, et al. 2013). Lastly, mean frame-wise displacement of both directions were regressed out
188 from a predicted environmental variable before it was entered into the prediction analysis.

189 **2.3.4. Structural MRI processing**

190 For the Day2day dataset, the MPRAGE anatomical MRI images were available for all the sessions.
191 Therefore, we used the anatomical images as a control condition for weather prediction. The analysis was
192 also performed in subject's native space. The anatomical images from all the sessions of a subject were
193 realigned and resliced to the image of the first session. Then each session's image was segmented
194 separately, and the segmented tissue probability maps of GM, WM, and CSF were obtained in the native
195 space. We defined GM, WM, and CSF masks as averaged probability greater than 0.5 in respective tissue
196 type. GM, WM, and CSF probability in their masks were extracted, respectively, to be used in the
197 prediction analysis.

198 **2.3.5. Phantom image processing**

199 For each session, the two images were realigned, and an averaged image was calculated. Because the
200 phantom was placed in the similar location, no cross-session registration was performed. We first
201 calculated correlations between daylight length and image values in every voxel, resulting in a correlation
202 image. Next, a cubic mask in the center of the image was defined. The signals within the mask were
203 extracted for the prediction analysis.

204 **2.4. Prediction analysis**

205 **2.4.1. Prediction analysis scheme**

206 The goal of the analysis is to estimate the prediction values of resting-state parameters on different
207 weather or meteorological parameters. The analysis was performed for each of the resting-state
208 parameters to predict each of the seven weather parameters. And we asked which weather parameters

209 could be better predicted by which resting-state parameters. The prediction analysis was all done in a
210 within-subject manner. Cross-validation was used to evaluate the prediction accuracy.

211 In addition to use these resting-state parameters, we also performed a series of control analysis to
212 use other potential confounding parameters to predict the environmental parameters. First, we used the
213 first fMRI image of each session after realignment to perform prediction analysis. Although the single
214 image still reflects BOLD effects, it may be more contributed by brain structural variations. Secondly, to
215 future rule out the structural contribution, we used segmented tissue probabilities of GM, WM, and CSF
216 from their respective tissue masks to perform prediction analysis. Thirdly, we also extracted raw image
217 values from the MPRAGE images in the three masks to serve as another control conditions. In addition, a
218 cuboid mask was also defined for each subject, which was located outside the brain. The raw image
219 values from the MPRAGE images from the air mask was used to control for baseline MRI signals.

220 Finally, since all of the above mentioned analyses indicated prediction values to predict environmental
221 parameters, especially daylight length, we further analyzed the quality assurance phantom data, and used
222 the signals in the agar phantom area to perform prediction analysis to predict daylight length.

223 **2.4.2. Machine learning regression analysis**

224 We used linear machine learning regression model to perform prediction analysis. The general form of
225 the prediction model is a linear regression model as the following:

$$226 \quad y = X \cdot \beta + \varepsilon$$

227 where y is a $1 \times n$ vector of a predicted weather parameter, X is a $n \times m$ matrix of a resting-state parameter,
228 β is the model parameters, and ε is the residual. N represents the number of observations, which in the
229 current analysis were the number of sessions for the longitudinal data and the number of subjects for the
230 cross-subject data. M represents the number of prediction variables, which could be the number of voxels
231 in the ALFF or ReHo maps (see Table 1) or the number of connections (13,366) in the connectivity
232 matrices. Here, m is much larger than n . Therefore, we used ridge regression to estimate the β

233 parameters. Briefly speaking, instead of trying to achieve the goal of minimizing the sum of square
234 means of the model prediction:

$$235 \quad \min_{\beta} \sum (y - X \cdot \beta)^2$$

236 Ridge regression adds one more regularized term:

$$237 \quad \min_{\beta} \sum (y - X \cdot \beta)^2 + \lambda \cdot \sum \|\beta\|^2$$

238 where λ represents the regularization parameter. The regularization term can constrain the sizes of beta
239 values, thus preventing over fitting of the model. In the current analysis, we used the MATLAB function
240 fitrlinear to perform the prediction analysis. There are other methods available, such as LASSO and
241 elastic net, but a recent study suggested that ridge regression and elastic net can yield similar prediction
242 accuracies while LASSO might perform worse in the scenario that the number of observations is much
243 smaller than the number of features (Cui and Gong 2018).

244 There are three steps in the prediction analysis, 1) tuning regularization parameter λ to find the
245 optimal λ (λ tuning), 2) training the model using the training dataset and the optimal λ to obtain a
246 prediction model β (model training), and 3) estimating prediction accuracy by calculating correlations
247 between predicted and actual values in a separate testing sample (cross-validation). Cross-validation was
248 used to make sure that the estimated prediction accuracy was independent from the training data.
249 Therefore, it is important for the within-subject and cross-subject prediction. For the between-subject
250 prediction, however, prediction accuracy was estimated from a separate subject; therefore no cross-
251 validation was required.

252 Because of the limited number of data in one fold (13 observations in the least case), 3-fold cross-
253 validation was adopted. We used a nested tuning strategy to optimize the parameter λ (Figure 1A) (Cui
254 and Gong 2018). Specifically, we first used 3-fold cross-validation to hold out one third of the data as an
255 independent testing dataset, and used the remaining two thirds of the data as training and parameter
256 tuning dataset. The data were first sorted according to the tested weather parameter, and the three folds

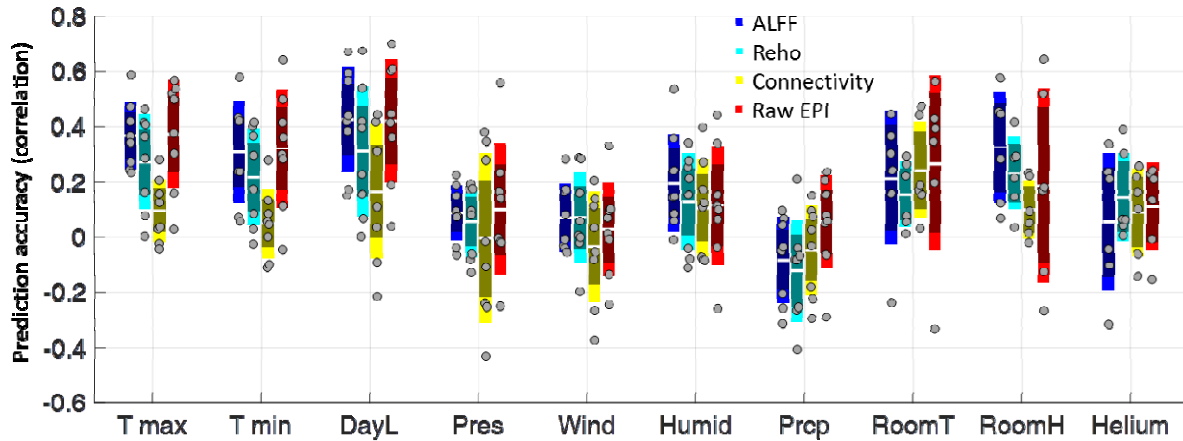
257 were defined as the 1st, 4th, 7th, ..., 2nd, 5th, 8th, ..., and 3rd, 6th, 9th, ... sessions of the data, respectively.
258 Within the two thirds training and parameter tuning dataset, we first performed a nested loop of 3-fold
259 analysis. Specifically, one third of the data were hold out, and the remaining of the two thirds of data
260 were used to train the regression model using a set of λ values, from 10^{-5} to 10^{-1} in the logarithmic scale
261 with a total 15 values. The one third of inner loop testing data was used to test the accuracy of the
262 prediction by calculating the correlation between predicted and actual weather parameter values. This
263 procedure was performed for three times for the three folds, and the mean accuracies were calculated for
264 each of the λ values. The λ value with the highest mean accuracy was used for the outer layer training
265 data to train the model. The model was then applied to the outer layer testing data to estimate prediction
266 accuracies. The three accuracy values from the 3 folds were averaged to represent an estimate of
267 accuracy for a subject.

268

269 **3. Results**

270 **3.1. Weather predictions using the resting-state data**

271 We first performed prediction analysis on different weather parameters using the ALFF maps, ReHo
272 maps, connectivity matrices, as well as single frame of EPI images as a control condition (Figure 1). In
273 general daylight length (DayL) and maximum and minimum environmental temperatures (T max and T
274 min) had prediction accuracies. And daylight length usually had the highest prediction accuracies. All
275 other environmental parameters had very low prediction accuracies. In terms of resting-state parameters,
276 the ALFF map usually had the highest prediction value. The averaged prediction accuracy of daylight
277 length was 0.42. However, using a single frame of EPI images could achieve comparable prediction
278 accuracies as the ALFF maps.



279

280 **Figure 1** Prediction accuracies (correlations) of the ALFF (amplitude of low-frequency fluctuations)

281 maps, ReHo (regional homogeneity) maps, and connectivity matrices on different weather parameters.

282 Each dot represents one subject's mean prediction accuracy. The center white line and inner dark bar

283 represent the mean and 95% confidence interval, respectively, and the outer light bar represents one

284 standard deviation to the mean. One sample t test was performed for each condition, and the red *

285 top represents statistical significant cross-subject effects at $p < 0.05$ after FDR (false discovery rate)

286 correction for all the 21 predictions. The figure was made by using notBoxPlot

287 (<https://github.com/raacampbell/notBoxPlot>).

288

289 It is still a question that whether the weather predictions using the resting-state parameters and

290 single EPI are based on similar or different information. Since the daylight length had the highest

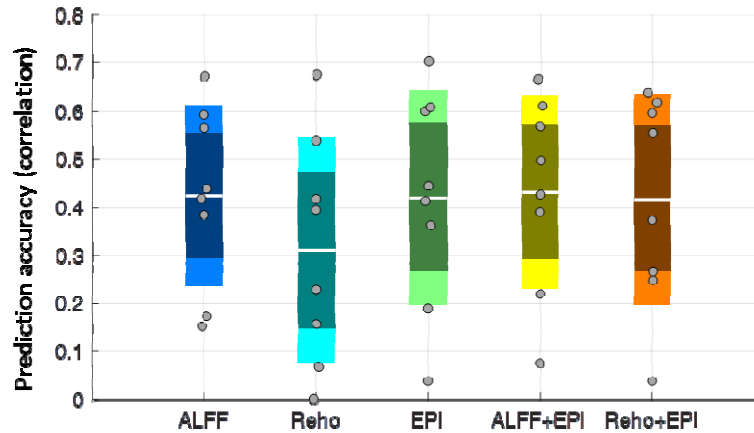
291 prediction accuracies, we focused on its prediction. We combined ALFF with EPI or ReHo with EPI to

292 predict daylight length to check whether combining the two modalities can boost the prediction accuracies.

293 Unfortunately, combining two modalities yielded very similar prediction accuracies as those using single

294 EPI images or ALFF images (Figure 2). Therefore, ALFF and ReHo did not convey more information

295 than a single frame EPI image to predict daylight length.



296

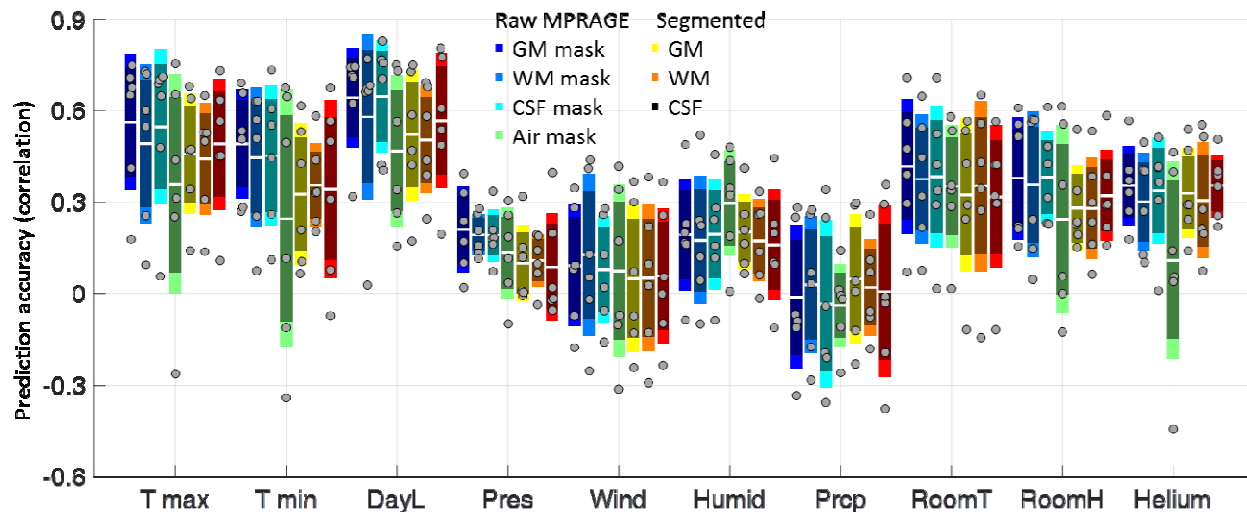
297 **Figure 2** Prediction accuracies to daylight length using the ALFF, Reho, raw EPI maps, and the
298 combination of EPI with ALFF or EPI.

299

300 3.2. Prediction analysis using the anatomical images

301 If single frame EPI images can predict weather parameter like daylight length, then the question becomes
302 whether the prediction is due to brain functional activity, structural information, or other factors. We
303 therefore performed the same prediction analysis using anatomical images, where were available in the
304 six subject in the Day2day dataset. We first performed prediction analysis using segmented GM, WM, or
305 CSF images using their respective tissue mask (Figure 3). The results showed very similar prediction
306 patterns for different environmental parameters as using resting-state parameters. That is, the daylight
307 length and environmental temperatures had the highest prediction accuracies. The prediction accuracies
308 using all the three tissue probability maps were above 0.5, which were higher than using any resting-state
309 parameters. However, what more interesting was that even higher prediction accuracies could be
310 achieved using raw anatomical images in different tissue masks. The prediction accuracies were higher
311 than 0.6 when using raw anatomical image signals in the GM or CSF masks. Finally, we defined a
312 cuboid mask outside the brain (see Figure 4A as an example), and used the raw MRI signals in the mask
313 to perform prediction analysis. Surprisingly, the analysis also showed similar pattern of prediction

314 accuracies. The prediction accuracy on daylight length using the air mask was 0.47, which was lower
315 than using all other anatomical parameters but still higher than using any of the resting-state parameters.

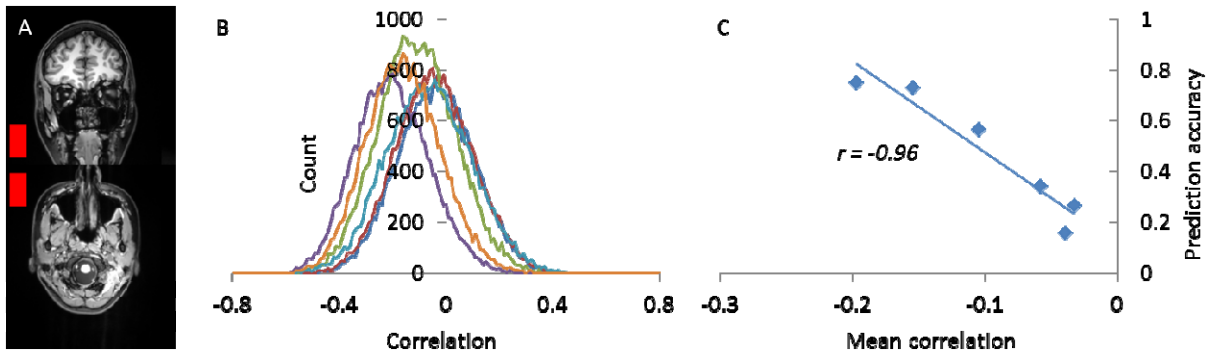


316

317 **Figure 3** Cross-subject predictions of weather parameters using the amplitude of low-frequency
318 fluctuations (ALFF) maps (top row), regional homogeneity (ReHo) maps (middle row), and connectivity
319 matrices (bottom row) from the six subjects of the Day2day dataset. T max and T min represent daily
320 maximum and minimum temperature, respectively.

321

322 To further explore the baseline MRI signals conveyed in the air mask, we calculated correlations
323 between MPRAGE signals and daylight length in all the voxels in the air mask in all the six subjects
324 (Figure 4B). There were small global effects of correlations between MPRAGE signals and daylight
325 length. The global effects of correlations were strongly correlated with the prediction accuracies across
326 subjects (Figure 4C), indicating that the global correlation is the driving information that gave rise to the
327 prediction accuracy.



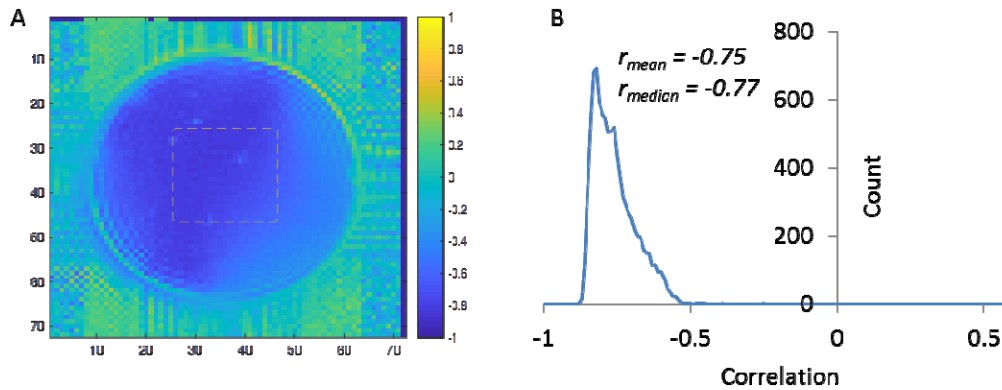
328

329 **Figure 4** A) An example of the air mask from one subject overlaid to the subject's mean anatomical
330 image. B) Histograms of the correlations between the MRI signals and daylight length of all the voxels in
331 the air mask. Each line with different colors represents one subject. C) There is an extremely high
332 negative correlation between the mean correlations in the air mask and the prediction accuracies of using
333 the air mask voxels to predict daylight length.

334

335 3.3. Control analysis using the phantom images

336 To further confirm the baseline signal changes, we analyzed the weekly quality control phantom data
337 around the same period of the Day2day project. We first calculated voxel-wise correlations between the
338 MRI signal and daylight length (Figure 5A). It clearly showed that in the phantom region, there were
339 high negative correlations. We defined a cubic mask in the center of the image, and the distribution of
340 correlations of all the voxels in the mask were plotted in Figure 5B. The mean and median correlation in
341 the cubic mask was -0.75 and -0.77 , respectively. Lastly, we performed the same kind of prediction
342 analysis using all the voxels in the mask to predict daylight length, and the cross-validated mean accuracy
343 was 0.72 .



344

345 **Figure 5** A) Voxel-wise Correlation maps between MRI image values and daylight length. The dash-line
346 contour indicates the location of the cubic mask. B) Histogram of the correlations between MRI image
347 values and daylight length in all the voxels in the cubic mask.

348

349

350 4. Discussion

351 By applying machine learning regression to longitudinal fMRI data that were scanned over months to
352 years, we demonstrated that we can predict environmental parameters, especially daylight length and
353 temperature, by using resting-state fMRI parameters. However, a series of control analysis showed that
354 using a single EPI image, segmented tissue images, and raw MPRAGE image signals in different tissue
355 masks and even in a mask outside the brain could all predict the environmental parameters. The resting-
356 state parameters did not add prediction values to single frame EPI images. These results indicated the
357 prediction of weather parameters, especially daylight length, cannot be explained as the weather effects
358 on brain functions. Rather, the prediction may reflect MRI scanner baseline signal variations that were
359 affected by environmental parameters. The analysis of the quality control phantom images supported our
360 speculation.

361 Among all the environmental parameters analyzed the daylight length had the highest prediction
362 accuracies. It's noteworthy that the prediction accuracies of daylight length were even higher than what
363 in the predictions of MRI scanner room temperature and humidity. Therefore, there may be some other

364 local factors that mediate the effects of daylight length on scanner stability. A study has shown that the
365 gaseous oxygen level in the magnet field can influence the MRI signals (Bates et al. 1995). It is
366 reasonable to speculate that the oxygen level in the scanner room may fluctuate across seasons due to
367 different ventilation conditions. In addition, the cooling systems of the scanner may be affected by either
368 electricity supply stability or cooling water temperature. Given that the MRI is such a sophisticated
369 machine, there may be other factors that mediate the association between daylight length and scanner
370 stability.

371 The current results highlighted the difficulty to study long term effects such as weather on brain
372 structures and functions using MRI. Consistent with two previous fMRI studies (Choe et al. 2015; Meyer
373 et al. 2016), we did find weather effects on fMRI measures. But we demonstrated that the weather effects
374 are likely due to the variations of MRI scanner stability. It is reasonable to speculate that MRI scanner
375 stability might contribute to the reported seasonal effects (Choe et al. 2015; Meyer et al. 2016). We also
376 showed that tissue probability measures of GM volumes may also be affected by the scanner stability, so
377 brain volumetric measures may also be affected by the scanner stability (Miller et al. 2015). The problem
378 may arise from the fact that most of the fMRI/MRI measures are relative measures. If absolute measures
379 can be used, e.g. using perfusion measures from arterial spin labeling (ASL) (Detre et al. 1992, 2012),
380 then the effects from scanner stability may be minimized. In addition, adjusting global signals maybe
381 helpful to minimize the scanner stability issue (Peelle et al. 2012; Yan, Craddock, et al. 2013). But at the
382 same time, the global signal was tightly coupled with neural activity (Scholvinck et al. 2010), and
383 adjusting global signals may cause statistical artifacts (Murphy et al. 2009; Saad et al. 2012).

384 The current analysis demonstrated that machine learning is a powerful method to be able to pick
385 up small effects. The phantom data showed the correlations between baseline MRI signals and daylight
386 length were about 0.7. Although the direct correlations of MRI signals and daylight length when
387 scanning human participants were blurred (Figure 4A), we could still achieve the similar level as the
388 prediction accuracies by using machine learning and cross-validation (Figure 3). Indeed, the cross-

389 validated prediction accuracies were between 0.6 and 0.7, which were very close to the correlation in the
390 phantom data. Machine learning methods have become more and more popular in studying brain-
391 behavior relationships (Finn et al. 2015; Cui and Gong 2018) and different mental disorders (Whelan et al.
392 2014). It is powerful on one hand, but on the other hand, careful control conditions are needed to rule out
393 possible confounding artifacts.

394 In conclusion, by applying machine learning on resting-state fMRI or structural MRI data, we
395 can predict several weather parameters, with the highest prediction accuracies to daylight length.
396 However, the predictions were not likely due to the weather effects on brain functions or structures, but
397 may due to the weather effects on the stability of MRI scanners. The data highlight the difficulty to use
398 fMRI/MRI data to study long term effects. And call for cautions to control for scanner stability when
399 studying long term effects.

400

401 **Acknowledgement**

402 We thank Dr. James Hyde for his insightful comments on an earlier version of this manuscript. This
403 analysis was funded by (US) National Institute of Health grants: R01AT009829 and R01DA038895.

404

405

406 **References:**

- 407 Bates S, Yetkin Z, Jesmanowicz A, Hyde JS, Bandettini PA, Estkowski L, Haughton VM. 1995. Artifacts
408 in functional magnetic resonance imaging from gaseous oxygen. *J Magn Reson Imaging*. 5:443–445.
- 409 Cedeño Laurent JG, Williams A, Oulhote Y, Zanobetti A, Allen JG, Spengler JD. 2018. Reduced
410 cognitive function during a heat wave among residents of non-air-conditioned buildings: An
411 observational study of young adults in the summer of 2016. *PLOS Med*. 15:e1002605.
- 412 Choe AS, Jones CK, Joel SE, Muschelli J, Belegu V, Caffo BS, Lindquist MA, Van Zijl PCM, Pekar JJ.
413 2015. Reproducibility and temporal structure in weekly resting-state fMRI over a period of 3.5 years.

- 414 PLoS One. 10.
- 415 Cui Z, Gong G. 2018. The effect of machine learning regression algorithms and sample size on
416 individualized behavioral prediction with functional connectivity features. *Neuroimage*.
- 417 Denissen JJA, Butalid L, Penke L, van Aken MAG. 2008. The effects of weather on daily mood: A
418 multilevel approach. *Emotion*. 8:662–667.
- 419 Detre JA, Leigh JS, Williams DS, Koretsky AP. 1992. Perfusion imaging. *Magn Reson Med*. 23:37–45.
- 420 Detre JA, Rao H, Wang DJJ, Chen YF, Wang Z. 2012. Applications of arterial spin labeled MRI in the
421 brain. *J Magn Reson Imaging*. 35:1026–1037.
- 422 Di X, Biswal BB. 2015. Characterizations of resting-state modulatory interactions in the human brain. *J*
423 *Neurophysiol*. 114:2785–2796.
- 424 Di X, Biswal BB. 2019. Toward Task Connectomics: Examining Whole-Brain Task Modulated
425 Connectivity in Different Task Domains. *Cereb Cortex*. 29:1572–1583.
- 426 Di X, Kannurpatti SS, Rypma B, Biswal BB. 2013. Calibrating BOLD fMRI Activations with
427 Neurovascular and Anatomical Constraints. *Cereb Cortex*. 23:255–263.
- 428 Dosenbach NUF, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, Nelson SM, Wig GS, Vogel AC,
429 Lessov-Schlaggar CN, Barnes KA, Dubis JW, Feczko E, Coalson RS, Pruett JR, Barch DM,
430 Petersen SE, Schlaggar BL. 2010. Prediction of individual brain maturity using fMRI. *Science*.
431 329:1358–1361.
- 432 Eisenberg DP, Kohn PD, Baller EB, Bronstein JA, Masdeu JC, Berman KF. 2010. Seasonal Effects on
433 Human Striatal Presynaptic Dopamine Synthesis. *J Neurosci*. 30:14691–14694.
- 434 Filevich E, Lisofsky N, Becker M, Butler O, Lochstet M, Martensson J, Wenger E, Lindenberger U,
435 Kühn S. 2017. Day2day: investigating daily variability of magnetic resonance imaging measures
436 over half a year. *BMC Neurosci*. 18:65.
- 437 Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT.
438 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain

- 439 connectivity. *Nat Neurosci.* 18:1664–1671.
- 440 Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R. 1996. Movement-related effects in fMRI
441 time-series. *Magn Reson Med.* 35:346–355.
- 442 Gillihan SJ, Detre JA, Farah MJ, Rao H. 2011. Neural Substrates Associated with Weather-Induced Mood
443 Variability: An Exploratory Study Using ASL Perfusion fMRI. *J Cogn Sci (Seoul).* 12:195–210.
- 444 Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C,
445 Gratton C, Sun H, Hampton JM, Coalson RS, Nguyen AL, McDermott KB, Shimony JS, Snyder AZ,
446 Schlaggar BL, Petersen SE, Nelson SM, Dosenbach NUF. 2017. Precision Functional Mapping of
447 Individual Human Brains. *Neuron.* 95:791–807.e7.
- 448 Hirshleifer D, Shumway T. 2003. Good Day Sunshine: Stock Returns and the Weather. *J Finance.*
449 58:1009–1032.
- 450 IJzerman H, Lindenberg S, Dalğar İ, Weissgerber SSC, Vergara RC, Cairo AH, Čolić M V., Dursun P,
451 Frankowska N, Hadi R, Hall CJ, Hong Y, Hu C-P, Joy-Gaba J, Lazarević D, Lazarević LB,
452 Parzuchowski M, Ratner KG, Rothman D, Sim S, Simão C, Song M, Stojilović D, Blomster JK,
453 Brito R, Hennecke M, Jaume-Guazzini F, Schubert TW, Schütz A, Seibt B, Zickfeld JH, Vazire S,
454 McCarthy R. 2018. The Human Penguin Project: Climate, Social Integration, and Core Body
455 Temperature. *Collabra Psychol.* 4.
- 456 Kaasinen V, Jokinen P, Joutsa J, Eskola O, Rinne JO. 2012. Seasonality of striatal dopamine synthesis
457 capacity in Parkinson’s disease. *Neurosci Lett.* 530:80–84.
- 458 Kalbitzer J, Erritzoe D, Holst KK, Nielsen FÅ, Marnier L, Lehel S, Arentzen T, Jernigan TL, Knudsen
459 GM. 2010. Seasonal Changes in Brain Serotonin Transporter Binding in Short Serotonin
460 Transporter Linked Polymorphic Region-Allele Carriers but Not in Long-Allele Homozygotes. *Biol*
461 *Psychiatry.* 67:1033–1039.
- 462 Keller MCM, Fredrickson BL, Ybarra O, Cote S, Johnson K, Mikels J, Conway A, Wager T, Côté S,
463 Johnson K, Mikels J, Conway A, Wager T. 2005. A warm heart and a clear head. The contingent

- 464 effects of weather on mood and cognition. *Psychol Sci.* 16:724–731.
- 465 Kurlansik SL, Ibay AD. 2012. Seasonal affective disorder. *Am Fam Physician.* 86:1037–1041.
- 466 Lim ASP, Gaiteri C, Yu L, Sohail S, Swardfager W, Tasaki S, Schneider JA, Paquet C, Stuss DT,
467 Masellis M, Black SE, Hugon J, Buchman AS, Barnes LL, Bennett DA, De Jager PL. 2018.
468 Seasonal plasticity of cognition and related biological measures in adults with and without
469 Alzheimer disease: Analysis of multiple cohorts. *PLOS Med.* 15:e1002647.
- 470 Mc Mahon B, Andersen SB, Madsen MK, Hjordt L V., Hageman I, Dam H, Svarer C, da Cunha-Bang S,
471 Baaré W, Madsen J, Hasholt L, Holst K, Frokjaer VG, Knudsen GM. 2016. Seasonal difference in
472 brain serotonin transporter binding predicts symptom severity in patients with seasonal affective
473 disorder. *Brain.* 139:1605–1614.
- 474 Meyer C, Muto V, Jaspas M, Kussé C, Lambot E, Chellappa SL, Degueldre C, Balteau E, Luxen A,
475 Middleton B, Archer SN, Collette F, Dijk D-J, Phillips C, Maquet P, Vandewalle G. 2016.
476 Seasonality in human cognitive brain responses. *Proc Natl Acad Sci.* 113:3066–3071.
- 477 Miller MA, Leckie RL, Donofry SD, Gianaros PJ, Erickson KI, Manuck SB, Roecklein KA. 2015.
478 Photoperiod is associated with hippocampal volume in a large community sample. *Hippocampus.*
479 25:534–543.
- 480 Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. 2009. The impact of global signal
481 regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage.*
482 44:893–905.
- 483 Peelle JE, Cusack R, Henson RNA. 2012. Adjusting for global effects in voxel-based morphometry: gray
484 matter decline in normal aging. *Neuroimage.* 60:1503–1516.
- 485 Poldrack RA, Laumann TO, Koyejo O, Gregory B, Hover A, Chen M-Y, Gorgolewski KJ, Luci J, Joo SJ,
486 Boyd RL, Hunicke-Smith S, Simpson ZB, Caven T, Sochat V, Shine JM, Gordon E, Snyder AZ,
487 Adeyemo B, Petersen SE, Glahn DC, Reese Mckay D, Curran JE, Göring HHH, Carless MA,
488 Blangero J, Dougherty R, Leemans A, Handwerker DA, Frick L, Marcotte EM, Mumford JA. 2015.

489 Long-term neural and physiological phenotyping of a single human. *Nat Commun.* 6:8885.

490 Praszak-Rieder N, Willeit M, Wilson AA, Houle S, Meyer JH. 2008. Seasonal Variation in Human
491 Brain Serotonin Transporter Binding. *Arch Gen Psychiatry.* 65:1072.

492 Saad ZS, Gotts SJ, Murphy K, Chen G, Jo HJ, Martin A, Cox RW. 2012. Trouble at rest: how correlation
493 patterns and group differences become distorted after global signal regression. *Brain Connect.* 2:25–
494 32.

495 Saunders EM. 1993. Stock prices and Wall Street weather. *Am Econ Rev.* 83:1337–1345.

496 Scholvinck ML, Maier A, Ye FQ, Duyn JH, Leopold DA. 2010. Neural basis of global resting-state fMRI
497 activity. *Proc Natl Acad Sci.* 107:10238–10243.

498 Song X-W, Dong Z-Y, Long X-Y, Li S-F, Zuo X-N, Zhu C-Z, He Y, Yan C-G, Zang Y-F. 2011. REST: a
499 toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS One.*
500 6:e25031.

501 Whelan R, Watts R, Orr CA, Althoff RR, Artiges E, Banaschewski T, Barker GJ, Bokde ALW, Büchel C,
502 Carvalho FM, Conrod PJ, Flor H, Fauth-Bühler M, Frouin V, Gallinat J, Gan G, Gowland P, Heinz
503 A, Ittermann B, Lawrence C, Mann K, Martinot J-L, Nees F, Ortiz N, Paillère-Martinot M-L, Paus T,
504 Pausova Z, Rietschel M, Robbins TW, Smolka MN, Ströhle A, Schumann G, Garavan H, Albrecht
505 L, Arroyo M, Aydin S, Bach C, Barbot A, Bricaud Z, Bromberg U, Bruehl R, Cattrell A, Czech K,
506 Dalley J, Desrivieres S, Fadai T, Fuchs B, Gollier Briand F, Head K, Heinrichs B, Heym N, Hübner
507 T, Ihlenfeld A, Ireland J, Ivanov N, Jia T, Jones J, Kepa A, Lanzerath D, Lathrop M, Lemaitre H,
508 Lüdemann K, Martinez-Medina L, Mignon X, Miranda R, Müller K, Nymberg C, Pentilla J, Poline
509 J-B, Poustka L, Rapp M, Ripke S, Rodehake S, Rogers J, Romanowski A, Ruggeri B, Schmäl C,
510 Schmidt D, Schneider S, Schroeder M, Schubert F, Sommer W, Spanagel R, Stacey D, Steiner S,
511 Stephens D, Strache N, Struve M, Tahmasebi A, Topper L, Vulser H, Walaszek B, Werts H,
512 Williams S, Peng Wong C, Yacubian J, Ziesch. V. 2014. Neuropsychosocial profiles of current and
513 future adolescent alcohol misusers. *Nature.* advance on.

- 514 Yan C-G, Craddock RC, Zuo X-N, Zang Y-F, Milham MP. 2013. Standardizing the intrinsic brain:
515 Towards robust measurement of inter-individual variation in 1000 functional connectomes.
516 Neuroimage. 80:246–262.
- 517 Yan CG, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, Li Q, Zuo XN, Castellanos FX,
518 Milham MP. 2013. A comprehensive assessment of regional variation in the impact of head
519 micromovements on functional connectomics. Neuroimage. 76:183–201.
- 520 Yuan R, Di X, Kim EH, Barik S, Rypma B, Biswal BB. 2013. Regional homogeneity of resting-state
521 fMRI contributes to both neurovascular and task activation variations. Magn Reson Imaging.
522 31:1492–1500.
- 523 Zang Y-F, He Y, Zhu C-Z, Cao Q-J, Sui M-Q, Liang M, Tian L-X, Jiang T-Z, Wang Y-F. 2007. Altered
524 baseline brain activity in children with ADHD revealed by resting-state functional MRI. Brain Dev.
525 29:83–91.
- 526 Zang Y, Jiang T, Lu Y, He Y, Tian L. 2004. Regional homogeneity approach to fMRI data analysis.
527 Neuroimage. 22:394–400.
- 528