

1 Ground-based observations of VLF waves as a proxy for satellite observations:  
2 Development of models including the influence of solar illumination and  
3 geomagnetic disturbance levels

4 Laura E. Simms<sup>1</sup>, Mark J. Engebretson<sup>1</sup>, Mark A. Clilverd<sup>2</sup>, and Craig J. Rodger<sup>3</sup>

5 <sup>1</sup> Department of Physics, Augsburg University, Minneapolis, MN, USA

6 <sup>2</sup> British Antarctic Survey (UKRI-NERC), Cambridge, UK

7 <sup>3</sup> Department of Physics, University of Otago, Dunedin, New Zealand

8

9 Key points:

10 Models accounting for transionospheric absorption and subionospheric attenuation improve satellite-  
11 ground VLF PSD correlations

12 Validation of these empirical models resulted in correlations between predicted and observed satellite  
13 VLF PSD of up to 0.764

14 Ground VLF receivers spaced around the Earth could provide longitudinal coverage of outer radiation  
15 belt chorus over  $\pm 45\text{-}75^\circ$  latitude

16

17

18 Abstract

19 Ground VLF observations have often been used to infer VLF activity in the magnetosphere, however,  
20 they are not an unbiased measure of activity at satellite altitudes due to transionospheric absorption  
21 and subionospheric attenuation. We propose several empirical models that control for these effects.  
22 VLF power spectral density (PSD) from the VLF/ELF Logger Experiment (VELOX, L=4.6, Halley, Antarctica)  
23 is used to predict DEMETER low Earth orbit VLF PSD. Validation correlations of these models are as high  
24 as 0.764, thus ground VLF receivers spaced around the Earth could provide complete coverage of outer  
25 radiation belt lower band chorus over the latitudinal limits of this model ( $\pm 45\text{-}75^\circ$ ). Correlations of four  
26 frequency bands (centered at 0.5 kHz, 1.0 kHz, 2.0 kHz, and 4.25 kHz) are compared. The simple linear  
27 correlation between ground and satellite VLF PSD in the 1.0 kHz channel was 0.606 (at dawn). A cubic  
28 model resulted in higher correlation (0.638). VLF penetration to the ground is reduced by ionospheric  
29 absorption during solar illumination and by disruption of ducting field lines during disturbed conditions.  
30 Subionospheric attenuation also reduces VLF observations from distant field lines. Addition of these  
31 covariates improved predictions. Both solar illumination and disturbed conditions reduced ground  
32 observation of VLF PSD, with higher power waves penetrating to the ground proportionately less than  
33 lower power waves. The effect of illumination in reducing wave penetration was more pronounced at  
34 higher frequency (4.25 kHz), with the effect at a mid-range frequency (2.0 kHz) falling between these  
35 two extremes.

36

37

38 Key words:

39 Ground VLF observations

40 Outer radiation belt VLF waves

41 Prediction models

42

43

44

45 1. Introduction

46

47 VLF chorus waves (very low frequency: 0.3-10 kHz discrete waves) are thought to play an important role  
48 in accelerating electrons to damaging relativistic speeds in the radiation belts, with waves in the lower  
49 band (0.1-0.5 of the fce (electron cyclotron frequency)) thought to be most effective (Horne and Thorne  
50 1998; Summers et al., 1998). Several studies have found correlations between lower band chorus and  
51 increased relativistic electron flux (Li et al., 2014; MacDonald et al., 2008; Meredith et al., 2002; Rodger  
52 et al., 2016; Simms et al., 2018a; 2018b; Smith et al., 2004). Hiss (100s of Hz-several kHz incoherent  
53 waves), on the other hand, is associated with relativistic electron precipitation (Hardman et al., 2015;

54 Hayosh et al., 2013; Lyons et al., 1972; Meredith et al., 2006; Summers et al., 2007; Tsurutani et al.,  
55 1975). Thus, any attempts to explain the levels of relativistic electron flux must consider these waves.

56 However, satellite VLF data, measured in the radiation belts, has not been readily available during much  
57 of the time over which radiation belt electron flux data has been collected. For this reason, statistical  
58 studies attempting to correlate VLF wave activity with radiation belt electron flux have often used  
59 ground-based observations of VLF waves (Simms et al., 2014; 2016; Smith et al., 2004) or proxies based  
60 on various measures of the electron population in the radiation belts (Li et al., 2014; MacDonald et al.,  
61 2008). Simultaneous observations of VLF chorus events from both satellite and ground stations suggest  
62 that ground-observed chorus ought to be a reasonable proxy for satellite observations (Demekhov et al.,  
63 2017; Martinez-Calderon et al., 2016; Nemeč et al., 2016; Titova et al., 2015). Case studies of particle  
64 microburst precipitation from the radiation belts also show an association with ground-observed VLF  
65 activity (Douma et al., 2018). However, in a statistical study, daily ground-observed VLF activity does not  
66 correlate well with electron flux at geosynchronous orbit (Simms et al., 2014; 2016), in comparison to  
67 the more robust correlations found between satellite-observed VLF waves and flux (Simms et al., 2018a).  
68 In this study, we study the relationship between ground-observed VLF power spectral density (PSD) at  
69 Halley, Antarctica and that observed by the DEMETER satellite. From this data, we hope to generate a  
70 better model for estimating VLF waves occurring in orbit from that observed on the ground.

71 Waves generated in the magnetosphere are ducted down field-aligned paths to the Halley station at  
72  $L \sim 4.6$ . (The  $L$  value is the distance in Earth radii at which a given magnetic field line crosses the Earth's  
73 magnetic equator.) Chorus is most likely to be observed during the dawn period (Golden et al., 2009).  
74 At the Halley, Antarctica VELOX ground station instrument, this dawn peak of chorus occurs at 9-12 UT  
75 (6-9 MLT) (Smith et al., 2010). The equatorial electron gyrofrequency ( $f_{ce}$ ) at  $L = 4.6$  is  $\sim 10$  kHz (Clilverd  
76 et al., 2012), thus lower band chorus that would propagate away from the equator at the geomagnetic  
77 latitude of Halley lies between 1 and 5 kHz ( $0.1-0.5f_{ce}$ ). The ducting of VLF waves to the ground is  
78 disrupted during ionospheric ionization due to collisions with irregularities (Lehtinen & Inan, 2009).  
79 Absorption due to ionization can occur both during geomagnetic disturbances due to increased auroral  
80 electrons (Ozaki et al., 2009) and during periods of solar illumination (Smith et al., 2010). Although  
81 these waves, once below the ionosphere, can travel quite far (at least up to 300 km --Ozaki et al., 2008),  
82 their spread from distant field lines is reduced by subionospheric attenuation (Challinor 1967; Smith and  
83 Jenkins 1998; Smith et al., 2010). Both absorption and attenuation are more influential during the day  
84 than at night and, therefore, also more influential during the summer months at Halley. The degree to  
85 which they act varies with frequency. Both absorption and attenuation act to reduce ground-observed  
86 VLF wave power at 1.0 kHz. Subionospheric attenuation has been found to peak in influence around 2-3  
87 kHz, with effects decreasing at higher frequencies (Challinor 1967; Figure 10.14 of Davies 1990).  
88 However, absorption during periods of solar illumination increases significantly at higher frequencies.  
89 This leads to terrestrial influences, such as sferics from lightning, dominating in ground observations  
90 above 10 kHz during the day because of the much higher ionospheric absorption in these higher  
91 frequencies (Smith et al., 2010), but even at lower frequencies, absorption can have a significant  
92 influence. Limiting ground VLF observations to dawn when chorus is seen and to the winter months  
93 when there is no solar illumination, could, therefore, result in better representation of VLF chorus waves  
94 in orbit. A previous study using only dawn observations in winter months from Halley, Antarctica  
95 resulted in a moderate improvement in correlation with electron flux compared to data over the entire  
96 year and the full 24 hour period (Simms et al., 2015). While limiting observations to the dawn period

97 allows sampling on a daily basis, limiting data collection to the winter months results in losing data for  
98 half the year. This can severely impact the ability to use ground data in studies. There has also been no  
99 direct means of assessing exactly how well the ground station observations represent VLF waves in the  
100 radiation belt where electron flux is measured.

101 VLF wave penetration to the ground is thought to be more efficient during quiet geomagnetic periods  
102 due to the availability of wave guiding structures and reduced ionospheric attenuation. Disturbed  
103 conditions result in the breakup of these structures and thus less efficient ducting of VLF waves to the  
104 ground (Golkowski et al., 2011). Therefore, ground-based VLF observations may be less reliable during  
105 the very periods when the VLF waves are most likely to be driving other geomagnetic processes such as  
106 electron enhancement and precipitation.

107 VLF observations from the DEMETER satellite provide an opportunity to establish whether ground-based  
108 VLF observations accurately represent the wave activity in orbit. In this study, we determine at which L  
109 shell and frequency band the satellite is best correlated with the ground observations as well as how this  
110 differs between dayside, nightside and dawn. Using Halley-observed VLF PSD as the dependent variable  
111 in multiple regression, we explore whether solar illumination (responsible for transionospheric  
112 absorption), longitudinal separation between satellite and ground (representing subionospheric  
113 attenuation), and geomagnetic disturbance level (leading to less efficient ducting and increased  
114 absorption) influence the penetration of satellite-observed VLF waves to the ground.

115 Although correlation analysis does not discriminate between the explanatory and predictor variable,  
116 regression analysis makes this distinction. While we use the first set of regression models to determine  
117 the influence of predictors on penetration of VLF waves to the ground, these cannot be used as proxy  
118 models to predict VLF waves at the altitude of the satellite from ground-observed data. To create  
119 predictive models we must reverse the explanatory and predictor VLF variables, using the satellite data  
120 as the dependent variable predicted by the ground data together with the covariates of distance,  
121 illumination, and geomagnetic disturbance. We validate these models using a portion of the data held  
122 in reserve.

123

124

## 125 2. Data and statistical methods

126 Satellite-observed VLF power spectral density (PSD) data ( $\log(\mu V^2/m^2/Hz)$ ) were obtained from the ICE  
127 (Instrument Champ Electrique) on the DEMETER satellite which was in Sun-synchronous orbit 2004-2010  
128 (Berthelier et al., 2006). (We use data from 2004-2007 as this overlaps with observations from the  
129 Halley ground station.) Observations from the same frequency bands as the Halley channels (0.5, 1.0,  
130 2.0, and 4.25 kHz) were averaged over each hour and categorized as from the dayside pass of the  
131 satellite (10:30 LT) or the nightside pass (22:30 LT). DEMETER was in low-Earth orbit, so most  
132 observations occurred over McIlwain L shells 2-4, with a lower number of observations at L shell 5. The  
133 operation of DEMETER caused no data to be collected at the highest latitudes, severely limiting the  
134 higher L-coverage. The low-Earth polar orbit resulted in limiting observations to roughly  $\pm 45-75^\circ$   
135 latitude over L shells 2-4.

136 For ground station data, we use the 4 VLF frequencies from the VELOX (VLF/ELF Logger Experiment) of  
137 Halley, Antarctica ( $L = 4.6$ ) centered at .5 kHz (width of .5 kHz), 1.0 kHz (width of 1.0 kHz), 2.0 kHz (width  
138 of 1.0 kHz), and 4.25 kHz (width of 1.5 kHz) (Smith et al., 2010). Data were hourly averaged. Isolated  
139 observations of  $>6$  mean log power spectral density ( $\log(10^{-33} \text{ T}^2 \text{ Hz}^{-1})$ ) were removed from the  
140 dataset. This corresponds to the upper limit of 60 dB shown in the plots of Smith et al. (2010). To  
141 compare to dayside and nightside passes of DEMETER, we use Halley data roughly 6 hours on either side  
142 of the satellite pass: 0600-1800 LT and 1800-0600 LT, respectively. This centered the satellite pass  
143 within the longitudinal range of Halley as the ground station passed under the satellite. Although the  
144 dayside satellite passes were near local noon and the nightside passes near local midnight, at Halley,  
145 both of these time periods may be illuminated (during Halley summer) or not illuminated (during Halley  
146 winter). The designation of dayside (noon) or nightside (midnight) refers to whether the Earth is  
147 oriented toward the sun or away from the sun, respectively. We also specifically model hourly-averaged  
148 dawn period data from Halley (0600-0900 MLT) when chorus is most likely to be observed (Smith et al.,  
149 2010).

150 Our initial analysis found the highest correlations with Halley ground observations were with DEMETER  
151 observations in L3 ( $L = 3.0-3.99$ ) and L4 ( $L = 4.0-4.99$ ). We use these L shell ranges (L3 and L4) in the  
152 further analyses. As the satellite was rarely exactly over the ground station, magnetic longitude (IGRF  
153 model) was used to calculate the longitudinal separation (in degrees) of DEMETER from Halley. Kp index  
154 data (where  $Kp > 2.3$  is considered disturbed geomagnetic conditions) was obtained from OMNIWeb.

155 Solar elevation calculations are summarized in Othman et al. (2018). The multiple regression analyses  
156 used are described in Neter et al. (1985). When comparing the effects of predictors on a common scale,  
157 standardized regression coefficients are reported from the multiple regressions. To produce  
158 standardized coefficients, variances of all variables in the model are standardized to 1. These  
159 coefficients then represent how many standard deviations the dependent variable will change when a  
160 particular predictor changes by one standard deviation. However, unstandardized coefficients are  
161 reported for the final predictive models to allow new predictions of DEMETER data from the ground  
162 Halley observations.

163 Interaction terms in the models were obtained by multiplying parameters. These interaction terms  
164 describe the difference response of the predicted variable to one explanatory variable when a second  
165 explanatory variable changes in value.

166 Quadratic and cubic terms were added to models to describe the change in the relationship between  
167 ground and satellite observations at varying levels of power spectral density. At low PSD ( $< 0.5 \log(10^{-33}$   
168  $\text{T}^2 \text{ Hz}^{-1})$  at Halley), the DEMETER satellite is better able to observe signals that are somewhat obscured  
169 below the noise floor limit of the Halley VELOX instrument. This may be due to lightning interference or  
170 VELOX instrument noise below this level. However, the ground station is still weakly picking up signal  
171 below this "noise floor" as there is still some relationship between the ground and satellite observed  
172 levels. For example, when ground observations are limited to below this 0.5 kHz threshold, the  
173 correlations in the 1.0 kHz channel at L3 ( $r=0.531$ ) and L4 ( $r=0.441$ ) between ground and satellite are still  
174 considerable. For this reason, we chose not to discard these observations, but to describe them. The  
175 slope of the relationship, however, changes considerably above the 0.5 kHz noise threshold. For this  
176 reason, a simple linear fit over the whole range is not the best model. We find that the addition of

177 quadratic and cubic terms to the regression allows a better fit, with the prediction line curving upward  
178 at higher VLF activity to show the changed relationship over this range.

179 Models predicting DEMETER VLF PSD from Halley observations were produced using years 2004, 2005,  
180 and 2007 as the training set. Year 2006 was used to test these models, by correlating observed  
181 DEMETER VLF PSD with that predicted by the Halley data. We fit a linear model predicting DEMETER  
182 data from Halley VLF observations, a cubic model (using linear, square, and cubic terms of Halley VLF),  
183 and a cubic model with covariates (solar illumination and Kp along with their interactions with Halley  
184 VLF). We present models both with and without longitudinal distance and hemisphere, the latter  
185 creating a more global model.

186 Model fits can be compared using  $R^2$  (coefficient of determination or prediction efficiency) which is the  
187 fraction of variation in the data explained by the model. However, for validation we calculated  
188 shrinkage by subtracting validation  $r^2$  (correlation between observations and predicted values) from the  
189  $R^2$  of the original regression model. This gave us an estimate of how well the model predicted satellite  
190 VLF PSD in a new dataset (Muller and Fetterman, 2002).

191 Statistical analyses were performed in IBM SPSS Statistics and MATLAB.

192

### 193 3. Results

194 Over the DEMETER L range from L2 to L5, in the 1.0 kHz band, VLF PSD dayside satellite observations  
195 correlate best with ground-based dawn observations (0600-900 MLT), with correlations ranging from  
196 0.44-0.61 depending on the L-shell (Figure 1). Overall dayside correlations (0600-1800 LT) were  
197 somewhat lower (0.33-0.46). Nightside observations (1800-0600 LT) showed even lower correlations  
198 (0.20-0.39). As expected, satellite data from L4 correlates well with observations from the ground  
199 station which lies at  $L \sim 4.6$ . However, the correlations of L3 DEMETER with Halley are all somewhat  
200 higher than the L4 correlations. Satellite and ground observations correlate less well in the 0.5, 2.0, and  
201 4.25 kHz bands. (At 0.5 kHz, L5, nightside, the bar is missing because the correlation was nearly zero.)

202

#### 203 3.1 Halley VELOX 1.0 kHz channel

204 We continue building our models with the 1.0 kHz L3 and L4 observations. Lines predicting VLF PSD  
205 levels observed by the Halley VELOX 1.0 kHz channel (ground station) from DEMETER L3 and L4 VLF PSD  
206 (satellite) are presented in Figure 2. We use least squares regression to fit a linear model:

$$207 \text{ Halley} = b_0 + b_1 \times \text{DEMETER} \quad (1)$$

208 and a cubic model:

$$209 \text{ Halley} = b_0 + b_1 \times \text{DEMETER} + b_2 \times \text{DEMETER}^2 + b_3 \times \text{DEMETER}^3 \quad (2)$$

210 The cubic terms capture some of the nonlinear relationship between ground and satellite observations.  
211 This allows including Halley VELOX measurements below the 0.5 kHz noise floor in the model.  
212 Correlation coefficients for both models are reported. While the correlation for the linear model is the

213 usual Pearson's  $r$ , the model correlation of the cubic model is the square root of the  $R^2$  (coefficient of  
214 determination).

215 The linear correlation between dayside Halley and DEMETER of 0.456 (L3, Figure 2a) and 0.444 (L4, 2d)  
216 are both improved if the observations are limited to the dawn period when chorus is most strongly seen.  
217 Correlations in the dawn period are 0.606 and 0.549 for L3 and L4, respectively (Figure 2b and e).  
218 Nightside correlations are not as high (0.386 and 0.361 for L3 (Figure 2c) and L4 (Figure 2f)). Cubic  
219 models fit the data somewhat better for all categories. For the dawn period, the cubic model  
220 correlation is raised to 0.637 (L3) and 0.597 (L4).

221 A correlation above 0.6 shows we have a reasonable empirical representation of the relationship  
222 between ground and satellite data. However, by including more physical processes we may be able to  
223 improve this proxy measure. As VLF wave occurrence is not a global phenomenon, satellite and ground  
224 station may see different localized activity when they are far apart. To correct for this possibility, we  
225 add longitudinal separation between satellite and ground station as well as the satellite hemisphere to  
226 the cubic model. Hemisphere is coded as +1 for south (i.e., the same as Halley) vs. -1 for north. In  
227 addition, there are likely to be factors that create an observation bias at the ground station. Solar  
228 illumination, due to increased ionospheric absorption, and disturbed conditions may both restrict the  
229 ducting of waves to the ground station. This would result in lower VLF activity seen at the ground vs. the  
230 satellite during summer months and periods of high geomagnetic activity. To correct for this, we add  
231 the sun's elevation and Kp to the models. However, it is possible that low VLF activity does not  
232 penetrate to the ground as effectively as high activity during periods of illumination or disturbance. This  
233 could lead to further bias in VLF observations during these periods. To study this, we also add  
234 interaction terms to the regression model. These are obtained by multiplying the explanatory factors  
235 (e.g., Illumination and Halley VLF PSD). A positive significant effect of this factor would indicate that  
236 higher VLF activity is predicted at DEMETER by one of these factors when the other is high. The full  
237 multi-factor model we test is:

$$\begin{aligned} Halley = & b_0 + b_1 \times DEMETER + b_2 \times DEMETER^2 + b_3 \times DEMETER^3 + b_4 \times Illumination \\ & + b_5 \times Illumination \times DEMETER + b_6 \times Longitude + b_7 \times Longitude \\ & \times DEMETER + b_8 \times Hemisphere + b_9 \times Kp + b_{10} \times Kp \\ & \times DEMETER \end{aligned} \quad (4)$$

238 The standardized regression coefficients in Figure 3 (dawn only; L3: Figure 3a, L4: Figure 3b) show that  
239 the most important factor is the linear component of satellite VLF PSD. However, the significant square  
240 term of VLF shows that the relationship becomes stronger at higher VLF activity.

241 The additional variables increase the dawn cubic model correlations to  $r = 0.659$  (L3),  $0.624$  (L4)  
242 (standardized regression coefficients of Figure 3 and Table 1). They also improve the correlation over  
243 the whole dayside (Table 2:  $r = 0.491$  (L3) and  $0.490$  (L4)) and the nightside (Table 3:  $0.445$  (L3) and  
244  $0.426$  (L4)). Analysis of residual errors (plotting residuals vs. predicted values and a normal probability  
245 plot) showed that residuals were both randomly and normally distributed. This is confirmation that this  
246 model fits the data reasonably well.

247 Although the addition of more variables offers only a moderate improvement in the fit of the model, it  
248 does provide information about the influence of these covariates. Solar illumination is associated with  
249 lower VLF activity seen at the ground in both L3 and L4 on both nightside and dayside (Table 2 and 3),

250 although this effect is seen most strongly during dawn (Table 1 and Figure 3ab). However, this may  
251 represent a seasonal effect in addition to representing a possible reduction in wave penetration to the  
252 ground station due to ionospheric attenuation. To further explore whether high illumination reduces  
253 the efficiency of the ducting of waves to the ground, we add an interaction term (Illumination  $\times$   
254 DEMETER VLF). In the regression model, this compares the slope of the relationship between satellite  
255 and ground VLF under conditions of high and low illumination. In the dawn period at both L3 and L4, the  
256 negative Illumination  $\times$  DEMETER VLF interaction term demonstrates that high illumination impedes the  
257 penetration of the highest VLF activity to the ground more than it impedes lower VLF activity. This is  
258 graphically described by the interaction plot (Figure 4a) where there is a stronger relationship (higher  
259 slope) between satellite and ground VLF (L3) at the lower 50% of illumination (observations below the  
260 median solar elevation of 5.9°). A smaller proportion of the dawn satellite VLF activity reaches the  
261 ground when both VLF activity and illumination are strong.

262 The distance of the satellite from the ground station can be measured by longitudinal separation  
263 between the two and by whether the satellite is in the same (southern) or different (northern)  
264 hemisphere from Halley. We hypothesized that this might account for some of the difference between  
265 ground and satellite VLF measurements attributable to subionospheric attenuation. Increased  
266 longitudinal separation between satellite and ground had no effect on VLF activity seen at the ground  
267 station in the dawn period at L3 (Figure 3a). This may only be because longitudinal distance was less  
268 variable during the dawn period as the satellite was passing over the ground station at about the same  
269 distance in every observation. On the dayside, longitudinal separation did lower the VLF PSD seen on  
270 the ground (Figure 3b). Longitudinal separation apparently increased the observed ground VLF PSD on  
271 the nightside (Figure 3c). It may be that the satellite, passing over near midnight, sees less VLF activity  
272 than the ground station if it is near dusk or dawn.

273 At lower frequency (1 kHz), on the nightside and at dawn, satellite hemisphere had no effect in the  
274 regression. DEMETER observed waves seen by the southern hemisphere ground station were at the  
275 same level even when the satellite was over the northern hemisphere. This is expected as the source of  
276 the VLF waves is likely near the geomagnetic equator and will propagate equally towards both  
277 hemispheres. However, on the dayside overall, when DEMETER was in the same hemisphere as the  
278 ground station, VLF activity on the ground was somewhat more highly correlated with satellite  
279 observations when the satellite was in the same (southern) hemisphere. This is somewhat unexpected,  
280 given that equatorially-produced VLF waves are assumed to propagate equally north or south of the  
281 equator. This suggests, instead, that there may be some inhomogeneity in wave propagation.

282 Periods of geomagnetic disturbance ( $K_p > 2.3$ ) resulted in higher VLF activity, with a stronger effect in the  
283 dawn period. The interaction term was negative on both dayside and nightside (Figure 3, Tables 1-3),  
284 with high  $K_p$  and high satellite VLF PSD resulting in lower ground-observed VLF PSD than would have  
285 been predicted by each of these factors individually (dayside: Figure 4b). This interaction shows the  
286 disruption of ducting efficiency to the surface during periods of high geomagnetic activity, as well as  
287 potentially increased D-region absorption due to energetic electron precipitation from the outer  
288 radiation belt (Neal et al., 2015).

289

290 3.2 Halley VELOX 4.25 kHz channel

291 The 4.25 kHz channel at Halley correlates less well with DEMETER observations than the 1.0 kHz channel  
292 (Figure 1). Overall, the relationship between satellite and ground VLF PSD is more linear. Quadratic and  
293 cubic terms are not as strong (Figure 5). This is due to the noise floor at this frequency being less of a  
294 factor. However, at dawn, relative to other factors at this frequency, solar illumination more strongly  
295 reduces the VLF PSD levels seen on the ground due to greater ionospheric absorption (Smith et al.,  
296 2010). On the nightside, as at the lower frequency, waves are more likely to be seen at greater  
297 longitudinal distance, but this is effect is not as strong. For reasons that are not understood, the  
298 response at higher frequency (4.25 kHz) to hemisphere was different. At this higher frequency, at dawn,  
299 VLF activity on the ground was more highly correlated with satellite observations when the satellite was  
300 in the same (southern) hemisphere. A similar analysis of the 3.0 kHz channel (not shown) showed a  
301 response to illumination and distance midrange between the 1.0 and 4.25 kHz results.

302

### 303 3.3 Use of ground data as a proxy for satellite VLF

304 Because of the scarcity of satellite VLF data (as described in the introduction), ground data has often  
305 been used as a proxy for VLF activity in orbit (as described in the introduction). While the above models  
306 describe how various factors affect the VLF wave penetration from satellite orbit to the ground station,  
307 we may be interested in the opposite question: how well VLF activity measured at the Halley ground  
308 station can be used to represent satellite activity. As we have shown above, the linear correlation  
309 between ground and satellite can be improved by using a cubic model and adding other variables. In  
310 this section we compare several models predicting satellite observations from ground observations.

311 To produce a predictive model of satellite VLF PSD, we reverse the predictor and response variable in  
312 the regression models. We now predict satellite (DEMETER) VLF activity with ground (Halley) VLF  
313 activity using three models: 1) simple correlation, 2) cubic regression, 3) cubic regression with the  
314 additional covariates of solar illumination and the the illumination×VLF interaction, and 4) cubic  
315 regression with solar illumination, the illumination×VLF interaction, Kp, and the Kp×VLF interaction. We  
316 do not use longitudinal distance or hemisphere because we wanted to make a more general prediction.  
317 We withhold Year 2006 as the test set and produce the models using the other years. We report the  
318 unstandardized coefficients and  $R^2$  (coefficient of determination or prediction efficiency) of these  
319 models (Dawn: Table 4, Dayside: Table 5, Nightside; Table 6). The relative influence of these predictors  
320 cannot be determined from the unstandardized coefficients, but they can be used to calculate  
321 predictions for novel data from the unscaled 2006 ground VLF data. These predictions are then  
322 correlated with the actual data observed at the satellite for these same observations.  $R^2$  (fraction of  
323 variation explained by the original model), validation correlations (correlation between observations in  
324 the test set and predictions from the models), and shrinkage (the reduction in predictive power in a test  
325 set) are also reported in the tables. The shrinkage in the dawn period models was low, indicating that  
326 these models predicted new observations relatively well. Some shrinkage statistics during day and  
327 nightside were negative due to the poorer fit of the models to the training set data.

328 Scatterplots of observed vs. predicted PSD values (dawn, L3, 1.0 kHz) give further indication of how good  
329 predictions from Halley are (Figure 6). Predictions from a simple linear model show a correlation of  
330 0.603 with observed values. However, this is not a particularly good model as can be seen by the scatter  
331 of points around a line showing the relationship between observed and predicted values (Figure 6a).  
332 This simple model does not allow for values much below 0 (satellite VLF PSD lies in the range -3 – 3

333  $\log(10^{-33} \text{ T}^2 \text{ Hz}^{-1})$ ). The cubic model is an improvement in correlation ( $r = 0.709$ ) but is unable to predict  
334 values below -1 or above 2 (Figure 6b). The addition of only Illumination and the Illumination×VLF  
335 interaction did not improve the prediction ability of the cubic model. The correlation between observed  
336 and predicted was only 0.707. (This model is given in the tables but is not in the figure). However, a  
337 cubic model with illumination, Kp, and their interactions with ground VLF gives an improved fit, with a  
338 correlation between predicted and observed values of 0.764 (Figure 6c). The scatter of observed vs.  
339 predicted points also falls more within the range of actual satellite VLF values. The final model (cubic  
340 with additional covariates of illumination, Kp, and their interactions with VLF) provides an approximate  
341 proxy for what a satellite would observe although it is not exact. Squaring the correlation coefficient of  
342 0.764 gives an  $r^2$  value of 0.584. This means the predicted values of the model explain 58.4% of the  
343 variation seen in the observations. Once again, residual error analysis showed the residuals were both  
344 randomly and normally distributed.

345

346

347

#### 348 4. Discussion

349 The highest correlations between ground and satellite VLF PSD are seen in the 1.0 kHz band over L2 to  
350 L4, but it is not a perfect one to one correspondence. While there is a statistically significant linear  
351 correlation (up to 0.435 on the dayside when the DEMETER satellite is at L3) it can be increased to 0.606  
352 if observations are limited to Halley dawn (UT 9-12), the period during which chorus (coherent VLF) is  
353 most likely seen at Halley (Smith et al., 2010). Using a cubic regression further increases the correlation  
354 ( $r = 0.637$ ) (Figure 2). Nightside correlations are lower than those seen on the dayside.

355 Although Halley is at L 4.6, DEMETER observations over L2 to L4 all correlate almost equally well with  
356 the ground observations. VLF waves are therefore not confined to a specific L shell in orbit and the  
357 Halley ground station would appear to pick up VLF activity from a wider range than its fixed position at  
358  $L=4.6$  would suggest. This is reasonable given the known efficiency of VLF propagation in the Earth-  
359 ionosphere waveguide, particularly equatorward of Halley where ice thickness is low.

360 VLF waves in the magnetosphere are only observed at ground stations if they are ducted down field-  
361 aligned paths. The efficiency of this ducting may be disrupted by solar illumination of the ionosphere  
362 (Smith et al., 2010) or during geomagnetically disturbed periods (Smith et al., 2010; Golkowski et al.,  
363 2011). For these reasons, ground data on its own may not be a reliable indicator of VLF activity in the  
364 magnetosphere. However, as both these processes are measurable, we built models adding solar  
365 degrees above the horizon and Kp as covariates in an attempt to improve the correlation between  
366 ground and satellite observations. Longitudinal separation between ground and satellite as well as the  
367 satellite hemisphere were also added to the models. These additions improved the correlations (up to  
368 0.659 at 1.0 kHz in the dawn period when satellite is at L3: Table 1).

369 Solar illumination increased transionospheric absorption and was therefore responsible for a reduction  
370 in ground VLF PSD relative to that measured at the satellite. This effect was most pronounced at dawn,  
371 however, there was a similar, if smaller, response to solar illumination on the nightside. This may be  
372 due to contamination of the nightside observations by Halley observations nearer to dawn or dusk. As

373 noted by others, we found that the reduction of VLF waves observed on the ground due to absorption  
374 by the sunlit ionosphere is a greater factor at higher frequency (4.25 kHz) (Challinor 1967, Smith and  
375 Jenkins 1998, Smith et al., 2010). However, absorption was not constant over the whole range of VLF  
376 values. On the dayside (including dawn), the highest VLF power was less likely to come through to the  
377 ground station when illumination was high. Thus, due to absorption, ground observations are not only  
378 lower relative to satellite observations, they are also not in constant proportion. This may be because  
379 high illumination reduces the distance over which VLF waves can propagate subionospherically. This, in  
380 turn, makes the reception of VLF waves at Halley more susceptible to local ionospheric absorption levels  
381 during storms, either F-region storm composition effects, or D-region electron precipitation effects. If  
382 VLF wave activity is higher during storms, the system becomes more susceptible to local changes in  
383 ionospheric absorption.

384 While geomagnetic disturbances ( $K_p > 2.3$ ) often lead to higher VLF activity (Smith et al., 2010), the  
385 disruption of field lines may reduce the amount of VLF activity seen at the ground as compared to that  
386 seen in orbit. As expected, we found higher  $K_p$  to be associated with more VLF activity, but the negative  
387 interaction term between  $K_p$  and VLF showed that high disturbance preferentially reduced the  
388 penetration of the most intense wave activity to the ground. Wave guiding structures appear to be  
389 more available in quiet conditions (Golkowski et al., 2011). In addition, increased ionospheric  
390 absorption during geomagnetic disturbances are likely to decrease the efficiency of the coupling  
391 between space and ground (Ozaki et al., 2009; Smith et al., 2010).

392 We measured subionospheric attenuation by including distance (either longitudinal or latitudinal)  
393 between ground station and satellite. We hypothesized that when the satellite was further from the  
394 ground station, attenuation would reduce the wave activity seen at the satellite relative to the ground.  
395 At greater longitudinal and latitudinal distance, at the lower frequency (1.0 kHz), this effect was seen  
396 during the dayside passes. More of the VLF activity seen at the satellite was observed on the ground  
397 station both when satellite and ground station were closer longitudinally and when they were in the  
398 same hemisphere. This was not the case at the higher frequencies. At 4.25 kHz, longitudinal distance  
399 was not a significant factor and the effect of hemisphere was reduced. At 3.0 kHz (not shown), the  
400 attenuation influence due to distance was midway between that observed for 1.0 and 4.25 kHz, both  
401 longitudinally and latitudinally. This difference in attenuation effect agrees with observation (Challinor,  
402 1967) that subionospheric attenuation peaks at about 2 kHz, then becomes less influential at higher and  
403 lower frequencies. The theoretical reasons for this are discussed by Wait (1957; 2013).

404 At dawn, we did see a hemisphere effect similar to that for the dayside overall, but there was no  
405 longitudinal distance effect. This may be only because the limited time period (UT 9-12) meant the  
406 satellite was in much the same longitudinal position over the ground station at every observation. On  
407 the nightside, no latitudinal (hemisphere) effect was seen. However, increased longitudinal separation  
408 between ground and satellite resulted in higher VLF power readings on the ground. This may be an  
409 artifact of the higher VLF power on the dayside. While the ground station may observe higher VLF levels  
410 when closer to the dawn or dusk of nightside, the DEMETER satellite, always nearer to midnight on the  
411 nightside pass would not. When the ground station is closer to the dawn or dusk, the satellite (still at  
412 midnight) would be at its farthest distance from Halley. Longitudinal distance would be at a maximum  
413 just as the ground station is closer to the dayside, making longitudinal distance appear to be a positive  
414 influence.

415

#### 416 4.1 Ground data as a proxy for satellite observations

417 The prediction efficiency (i.e., coefficient of determination or  $R^2$ ) indicates how closely the data lie along  
418 the fitted regression line. It does not provide information on how well the model predicts new  
419 observations. For this reason, we perform validation tests of models, withholding year 2006 data as the  
420 test set. Although we report the  $R^2$  in the tables, the more important statistic is the correlation between  
421 observations and predictions in the test set.

422 Linear models predicting DEMETER VLF PSD from Halley ground data resulted in reasonable correlations  
423 between observed and predicted data (up to 0.603 in the dawn period). However, a cubic model  
424 provides a better fit to the test set (up to 0.709) and the addition of covariates not only improves  
425 correlation between observed DEMETER VLF PSD and that predicted by the model ( $r = 0.764$ ), it also  
426 results in a spread of predicted values that covers more of the natural range of DEMETER VLF  
427 observations. The added covariates of solar illumination and the illumination $\times$ HalleyVLF interaction  
428 account for absorption of VLF waves by the ionosphere and the tendency of higher power VLF to be  
429 preferentially absorbed.  $K_p$  as a covariate accounts for the higher VLF power seen during disturbed  
430 conditions, but, of more interest to the model, the  $K_p \times$ HalleyVLF interaction accounts for the reduced  
431 penetrance of VLF waves to the ground station due to disruption of ducting field lines during  
432 geomagnetic disturbances. (We did not add longitudinal distance or hemisphere to the model in order  
433 to make a more generalized prediction.)

434 While for ULF (ultralow frequency) wave power,  $K_p$  on its own appears to be a poor proxy (Murphy et  
435 al., 2016), VLF activity shows a correlation with  $K_p$  (Smith et al., 2010).  $K_p$  used as a covariate in a proxy  
436 model describing satellite VLF PSD from ground VLF data can improve predictions. However, if a VLF  
437 proxy was needed for a study designed to determine the effect of  $K_p$  on VLF waves, the cubic proxy  
438 model without covariates could be used with some loss of predictive ability.

439 As pitch angle scattering by chorus waves is a dominant driver of electron precipitation into the  
440 atmosphere, chorus wave amplitudes have also been inferred from low-altitude electron measurements  
441 made by POES (Polar Orbiting Environmental Satellites) (Li et al., 2013). In this study, predictions from a  
442 model based on electron pitch angle distributions of POES electron data correlated well during  
443 conjunction events with Van Allen Probes chorus observations, with  $r = 0.60$  over a 4 month period in  
444 2012. This is somewhat lower than our validation correlations using dawn or dayside Halley VELOX  
445 observations to predict satellite VLF activity ( $r = 0.764$  and  $0.682$ , respectively).

446 Our results suggest that ground VLF receivers spaced around the Earth could provide complete  
447 longitudinal (MLT) coverage of the satellite environment for lower band chorus in the outer radiation  
448 belt. Our models are limited to the  $\pm 45$ - $75^\circ$  latitudinal range where the DEMETER satellite observed L2-  
449 4. However, this is similar to the latitudinal range at these L shells of other low-earth polar orbit  
450 satellites such as POES. This would make extension of these results possible and could provide a  
451 cheaper alternative to the replacement of the POES satellites for VLF wave observation in these  
452 latitudes. It is important to note, however, that these models do not extend to the  $\pm 15$ - $30^\circ$  latitude  
453 range where other processes such as Landau damping and LHR reflection may limit the propagation of  
454 VLF waves both to the ground and to low-Earth orbit (Mourenas et al., 2012). A similar regression

455 model might be built to model VLF waves in the outer radiation belt from ground-based data at lower  
456 latitudes if satellite VLF observations were available at these latitudes.

457

458

## 459 5. Conclusions

460 1. Ground VLF power spectral density (PSD) observations (Halley,  $L \sim 4.6$ ) are not an unbiased  
461 measure of VLF PSD at satellite altitude (DEMETER).

462 2. Although there is a reasonable linear correlation between the two measures ( $r = 0.606$  during  
463 the dawn period at Halley, at 1.0 kHz and L3 at DEMETER), this correlation can be improved by  
464 correcting for transionospheric absorption during high solar illumination and by accounting for  
465 disruption of ducting processes along the field lines during geomagnetic disturbances ( $K_p > 2.3$ ).  
466 Adding interaction terms with these covariates also corrected the bias against penetration of  
467 high power VLF waves to the ground during conditions of high solar illumination and high  
468 geomagnetic disturbance.

469 3. A full cubic model with added covariates and interactions resulted in a correlation of 0.659  
470 with satellite VLF PSD.

471 4. A separate model (using a training set) predicting satellite VLF PSD with ground data and the  
472 covariates successfully predicted a withheld test set, with a correlation between test set  
473 observations and predictions of 0.764 (dawn, L3, 1.0 kHz).

474 5. Our results suggest that ground VLF receivers spaced around the Earth could provide  
475 complete longitudinal (MLT) coverage of the satellite environment for lower band chorus in the  
476 outer radiation belt. Although the models presented here are limited in latitudinal range ( $\pm 45$ -  
477  $75^\circ$ ) due to the DEMETER orbit, further models could be built covering the lower latitudes if  
478 satellite VLF wave data were available.

479

480

481

482

483

## 484 Acknowledgements

485 We thank R. Gamble for preparing and J.-J. Berthelier for providing DEMETER ICE data. All DEMETER  
486 data are now available at the CDPP (Centre de données de la Physique des Plasmas) website:  
487 [cdpp416archive.cnes.fr](http://cdpp416archive.cnes.fr).

488 Halley VELOX data from BAS are available at

489 <https://data.bas.ac.uk/metadata.php?id=GB/NERC/BAS/AEDC/00055>.  $K_p$  values were obtained from  
490 Goddard Space Flight Center Space Physics Data Facility at the OMNIWeb data website

491 ([http://omniweb.gsfc.nasa.gov/html/ow\\_data.html](http://omniweb.gsfc.nasa.gov/html/ow_data.html)). M. A. Clilverd was supported by the Natural  
492 Environmental Research Council grant NE/J008125/1. Work at Augsburg University was supported by  
493 NSF grants PLR-1341493 and AGS- 1651263.

494  
495

496

497

498

499 Literature Cited

500 Berthelier, J.J., M. Godefroy, F. Leblanc, M. Malingre, M. Menvielle, D. Lagoutte, J.Y. Brochet, F. Colin, F.  
501 Elie, C. Legendre, P. Zamora, D. Benoist, Y. Chapuis, J. Artru, R. Pfaff (2006), ICE, the electric field  
502 experiment on DEMETER, *Planetary and Space Science* 54, 456–471, doi:10.1016/j.pss.2005.10.016

503 Challinor, R. A. (1967), The phase velocity and attenuation of audio-frequency electro-magnetic waves  
504 from simultaneous observations of atmospheric at two spaced stations. *Journal of Atmospheric and*  
505 *Terrestrial Physics* 29, 803-810.

506

507 Clilverd, M. A., C. J. Rodger, D. Danskin, M. E. Usanova, T. Raita, T. Ulich, and E. L. Spanswick (2012),  
508 Energetic particle injection, acceleration, and loss during the geomagnetic disturbances which upset  
509 Galaxy 15, *Journal of Geophysical Research*, 117, A12213, doi:10.1029/2012JA018175

510 Davies, K. (1990), *Ionospheric Radio*, IEE electromagnetic waves series, 31, P. Peregrinus, London, UK

511 Demekhov, A. G., Manninen, J., Santolík, O., & Titova, E. E. (2017). Conjugate ground-spacecraft  
512 observations of VLF chorus elements. *Geophysical Research Letters*, 44, 11735-11744,  
513 doi.org/10.1002/2017GL076139

514 Douma, E., Rodger, C. J., Clilverd, M. A., Hendry, A. T., Engebretson, M. J., & Lessard, M. R. (2018).  
515 Comparison of relativistic microburst activity seen by SAMPEX with ground-based wave measurements  
516 at Halley, Antarctica. *Journal of Geophysical Research: Space Physics*, 123, 1279–1294.  
517 <https://doi.org/10.1002/2017JA024754>

518 Golden, D. I., M. Spasojevic, and U. S. Inan (2009), Diurnal dependence of ELF/VLF hiss and its relation to  
519 chorus at L = 2.4, *Journal of Geophysical Research*, 114, A05212, doi:10.1029/2008JA013946

520 Gołkowski, M., M. B. Cohen, D. L. Carpenter, and U. S. Inan (2011), On the occurrence of ground  
521 observations of ELF/VLF magnetospheric amplification induced by the HAARP facility, *Journal of*  
522 *Geophysical Research*, 116, A04208, doi:10.1029/2010JA016261

523 Hardman, R., M. A. Clilverd, C. J. Rodger, J. B. Brundell, R. Duthie, R. H. Holzworth, I. R. Mann, D. K.  
524 Milling, and E. Macusova (2015), A case study of electron precipitation fluxes due to plasmaspheric hiss,  
525 *Journal of Geophysical Research Space Physics*, 120,6736–6748, doi:10.1002/2015JA021429

526 Hayosh, M., D. L. Pasmanik, A. G. Demekhov, O. Santolik, M. Parrot, and E. E. Titova (2013),  
527 Simultaneous observations of quasi-periodic ELF/VLF wave emissions and electron precipitation by

528 DEMETER satellite: A case study, *Journal of Geophysical Research Space Physics*, 118, 4523–4533,  
529 doi:10.1002/jgra.50179

530 Horne, R. B. and R. M. Thorne (1998), Potential waves for relativistic electron scattering and stochastic  
531 acceleration during magnetic storms, *Geophysical Research Letters*, 25, 3011–3014,  
532 doi: 10.1029/98gl01002

533 Lehtinen, N. G., and U. S. Inan (2009), Full-wave modeling of transionospheric propagation of VLF  
534 waves, *Geophys. Res. Lett.*, 36, L03104, doi:10.1029/2008GL036535

535 Li, W., B. Ni, R. M. Thorne, J. Bortnik, J. C. Green, C. A. Kletzing, W. S. Kurth, G. B. Hospodarsky (2013),  
536 Constructing the global distribution of chorus wave intensity using measurements of electrons by the  
537 POES satellites and waves by the Van Allen Probes, *Geophysical Research Letters*, 40, 4526-4532,  
538 doi:10.1002/grl.50920

539 Li, W., R. M. Thorne, Q. Ma, B. Ni, J. Bortnik, D. N. Baker, H. E. Spence, G. D. Reeves, S. G. Kanekal, J. C.  
540 Green, C. A. Kletzing, W. S. Kurth, G. B. Hospodarsky, J. B. Blake, J. F. Fennell, and S. G. Claudepierre  
541 (2014), Radiation belt electron acceleration by chorus waves during the 17 March 2013 storm, *Journal of*  
542 *Geophysical Research Space Physics*, 119, 4681-4693, doi:10.1002/2014JA019945

543 Lyons, L.R., R.M. Thorne, and C.F. Kennel (1972), Pitch-Angle diffusion of radiation belt electrons within  
544 the plasmasphere, *Journal of Geophysical Research*, 77(19), 3455-3474, doi:10.1029/JA077i019p03455

545 MacDonald, E.A., M.H. Denton, M.F. Thomsen, and S.P. Gary (2008), Superposed epoch analysis of a  
546 whistler instability criterion at geosynchronous orbit during geomagnetic storms, *J. Atmos. Solar-Terr.*  
547 *Phys.*, doi:10.1016/j.jastp.2008.03.021

548 Martinez-Calderon, C., Shiokawa, K., Miyoshi, Y., Keika, K., Ozaki, M., Schofield, I.,...Kurth, W. S. (2016).  
549 ELF/VLF wave propagation at subauroral latitudes: Conjugate observation between the ground and Van  
550 Allen Probes A. *Journal of Geophysical Research: Space Physics*, 121, 5384–5393.  
551 <https://doi.org/10.1002/2015JA022264>

552 Meredith, N. P., Horne, R. B., & Anderson, R. R. (2001). Substorm dependence of chorus amplitudes:  
553 Implications for the acceleration of electrons to relativistic energies. *Journal of Geophysical Research*,  
554 106(A7), 13,165–13,178. <https://doi.org/10.1029/2000JA900156>  
555

556 Meredith, N. P., R. B. Horne, R. H. A. Iles, R. M. Thorne, D. Heynderickx, and R. R. Anderson (2002), Outer  
557 zone relativistic electron acceleration associated with substorm-enhanced whistler mode chorus,  
558 *J. Geophys. Res.*, 107(A7), 1144, 10.1029/2001JA900146

559 Meredith, N. P., R. B. Horne, S. A. Glauert, R. M. Thorne, D. Summers, J. M. Albert, and R. R. Anderson  
560 (2006), Energetic outer zone electron loss timescales during low geomagnetic activity, *Journal of*  
561 *Geophysical Research*, 111, A05212, doi:10.1029/2005JA011516

562 Mourenas, D., A. V. Artemyev, J.-F. Ripoll, O. V. Agapitov, and V. V. Krasnoselskikh (2012), Timescales for  
563 electron quasi-linear diffusion by parallel and oblique lower-band chorus waves, *Journal of Geophysical*  
564 *Research*, 117, A06234, doi:10.1029/2012JA017717

565 Muller, K.E. and B.A. Fetterman (2002), *Regression and ANOVA: An Integrated Approach Using SAS*  
566 *Software*, SAS Institute, Inc., Cary, NC, USA, 575 pp.

567 Murphy, K. R., I. R. Mann, I. J. Rae, D. G. Sibeck, and C. E. J. Watt (2016), Accurately characterizing the  
568 importance of wave-particle interactions in radiation belt dynamics: The pitfalls of statistical wave  
569 representations, *Journal of Geophysical Research Space Physics*, 121, 7895–7899, doi:10.1002/  
570 2016JA022618

571 Neal, J. J., C. J. Rodger, M. A. Clilverd, N. R. Thomson, T. Raita, and Th. Ulich (2015), Long-term  
572 determination of energetic electron precipitation into the atmosphere from AARDDVARK  
573 subionospheric VLF observations, *J. Geophys. Res.*, 120, 2194–2211, doi:10.1002/2014JA020689

574 Němec, F., Bezděková, B., Manninen, J., Parrot, M., Santolík, O., Hayosh, M., & Turunen, T. (2016).  
575 Conjugate observations of a remarkable quasiperiodic event by the low-altitude DEMETER spacecraft  
576 and ground-based instruments. *Journal of Geophysical Research: Space Physics*, 121, 8790–8803,  
577 doi:10.1002/2016JA022968

578 Neter, J., W. Wasserman, and M. H. Kutner (1985), *Applied Linear Statistical Models*, Richard D. Irwin,  
579 Inc., Homewood, Ill.

580 Othman, A. Ben, K. Belkilani, & M. Besbes (2018), Global solar radiation on tilted surfaces in Tunisia:  
581 Measurement, estimation and gained energy assessments, *Energy Reports* 4, 101–109,  
582 <https://doi.org/10.1016/j.egy.2017.10.003>

583 Ozaki, M., S. Yagitani, I. Nagano, Y. Hata, H. Yamagishi, N. Sato, and A. Kadokura (2008), Localization of  
584 VLF ionospheric exit point by comparison of multipoint ground-based observation with full-wave  
585 analysis, *Polar Science* 2, 237–249, doi:10.1016/j.polar.2008.09.001

586 Ozaki, M., S. Yagitani, I. Nagano, H. Yamagishi, and N. Sato (2009), Estimation of enhanced electron  
587 density in the lower ionosphere using correlation between natural VLF emission intensity and CNA,  
588 *Antarctic Record* 53(2):123-135, doi: 10.15094/00009496

589 Rodger C.J., K. Cresswell-Moorcock, & M.A. Clilverd (2016), Nature's Grand Experiment: Linkage  
590 between magnetospheric convection and the radiation belts, *Journal of Geophysical Research*, 121, 171-  
591 189, doi:10.1002/2015JA021537

592 Simms, L. E., M. J. Engebretson, A.J. Smith, M. Clilverd, V.A. Pilipenko, and G.D. Reeves (2014), Prediction  
593 of relativistic electron flux following storms at geostationary orbit: multiple regression analysis, *Journal*  
594 *of Geophysical Research Space Physics*, 119, 7297-7318, doi: 10.1002/2014JA019955

595 Simms, L. E., M. J. Engebretson, A.J. Smith, M. Clilverd, V.A. Pilipenko, G.D. Reeves, (2015), Analysis of  
596 the effectiveness of ground-based VLF wave observations for predicting or nowcasting relativistic  
597 electron flux at geostationary orbit, *Journal of Geophysical Research Space Physics*, 120, 2052-2060,  
598 doi:10.1002/2014JA020337

599 Simms, L. E., M. J. Engebretson, V. Pilipenko, G. D. Reeves, and M. Clilverd (2016), Empirical predictive  
600 models of daily relativistic electron flux at geostationary orbit: Multiple regression analysis, *Journal of*  
601 *Geophysical Research Space Physics*, 121, 3181–3197, doi:10.1002/2016JA022414

602 Simms, L., Engebretson, M., Clilverd, M., Rodger, C., Lessard, M., Gjerloev, J., & Reeves, G. (2018a). A  
603 distributed lag autoregressive model of geostationary relativistic electron fluxes: Comparing the

604 influences of waves, seed and source electrons, and solar wind inputs. *Journal of Geophysical Research: Space Physics*, 123, 3646–3671, doi:10.1002/2017JA025002

605

606 Simms, L.E., M. J. Engebretson, M. A. Clilverd, C. J. Rodger, M. R. Lessard, and G. D. Reeves (2018b),  
607 Nonlinear and synergistic effects of ULF Pc5, VLF chorus, and EMIC waves on relativistic electron flux at  
608 geosynchronous orbit. *Journal of Geophysical Research: Space Physics* doi:10.1029/2017JA025003

609 Smith, A.J. and P.J. Jenkins (1998), A survey of natural electromagnetic noise in the frequency range  $f =$   
610 1-10 kHz at Halley station, Antarctica: 1. Radio atmospherics from lightning, *Journal of Atmospheric and*  
611 *Solar-Terrestrial Physics*, 60 (2), 263-277.

612 Smith, A.J., N. P. Meredith, and T.P. O'Brien (2004), Differences in ground-observed chorus in  
613 geomagnetic storms with and without enhanced relativistic electron fluxes, *Journal of Geophysical*  
614 *Research*, 109, A11204, doi:10.1029/2004JA010491

615 Smith, A.J., R.B. Horne, and N.P. Meredith (2010), The statistics of natural ELF/VLF waves derived from a  
616 long continuous set of ground based observations at high latitude, *Journal of Atmospheric and Solar-*  
617 *Terrestrial Physics*, 72, 463-475.

618 Summers, D., R. M. Thorne, and F. Xiao (1998), Relativistic theory of wave-particle resonant diffusion  
619 with application to electron acceleration in the magnetosphere, *Journal of Geophysical Research*, 103,  
620 20,487, doi:10.1029/98JA01740

621 Summers, D., B. Ni, and N. P. Meredith (2007), Timescales for radiation belt electron acceleration and  
622 loss due to resonant wave-particle interactions: 2. Evaluation for VLF chorus, ELF hiss, and  
623 electromagnetic ion cyclotron waves, *Journal of Geophysical Research*, 112, A04207,  
624 doi:10.1029/2006JA011993

625 Titova, E. E., Kozelov, B. V., Demekhov, A. G., Manninen, J., Santolík, O., Kletzing, C. A., & Reeves, G.  
626 (2015), Identification of the source of quasiperiodic VLF emissions using ground-based and Van Allen  
627 Probes satellite observations. *Geophysical Research Letters*, 42, 6137–6145, doi:10.1002/2015GL064911

628 Tsurutani, B.T., E.J. Smith, and R.M. Thorne (1975), Electromagnetic hiss and relativistic electron losses  
629 in the inner zone, *Journal of Geophysical Research*, 80(4), 600-607

630 Tsurutani, B. T., & Smith, E. J. (1977), Two types of magnetospheric ELF chorus and their substorm  
631 dependences. *Journal of Geophysical Research*, 82(32), 5112–5128, doi:10.1029/JA082i032p05112

632

633 Wait, J.R. (1957), The Attenuation vs Frequency Characteristics of VLF Radio Waves, *Proceedings of the*  
634 *IRE*, 45(6), 768-771, doi: 10.1109/JRPROC.1957.278470

635

636 Wait, J.R. (2013), *Electromagnetic Waves in Stratified Media: Revised Edition Including Supplemented*  
637 *Material*. Pergamon Press, Oxford, UK

638

639

640



642

643 Table 1. Dawn models with standardized regression coefficients predicting ground VLF (1.0 kHz and 4.25 kHz) from satellite observations.

644 Prediction efficiency ( $R^2$ , fraction of variation in data explained by the regression model) and correlations ( $r$ ) are given. \*: statistically significant  
 645 coefficient ( $p < 0.05$ ).

Frequency	L shell	Model									$R^2$	$r$
		DEMETER VLF	DEMETER VLF <sup>2</sup>	DEMETER VLF <sup>3</sup>	Illumination	IllumXVLF	Longitude	Latitude	Kp	KpXVLF		
1 kHz	L3	0.606*									0.368	0.606
		0.540*	0.212*	-0.006							0.407	0.637
		0.671*	0.237*	-0.014	-0.116*	-0.088*	-0.008	0.026	0.135*	-0.209*	0.434	0.659
	L4	0.550*									0.302	0.549
		0.326*	0.204*	0.179*							0.357	0.597
		0.369*	0.186*	0.145*	-0.117*	-0.067*	-0.063	0.101*	0.192*	-0.097*	0.390	0.624
4.25 kHz	L3	0.352*									0.124	0.352
		0.415*	-0.075*	-0.097*							0.131	0.362
		0.681*	-0.038	-0.156*	-0.315*	-0.215	-0.030	0.113*	0.087*	-0.182*	0.215	0.463
	L4	0.305*									0.093	0.305
		0.246*	-0.013	0.059							0.094	0.306
		0.485*	0.025	-0.030	-0.323*	-0.156*	-0.085*	0.136*	0.113*	-0.136*	0.187	0.432

646

647 Table 2. Dayside models with standardized regression coefficients predicting ground VLF (1.0 kHz) from satellite data. Prediction efficiency ( $R^2$ ,  
 648 fraction of variation in data explained by the regression model) and correlations ( $r$ ) are given. \*: statistically significant coefficient ( $p < 0.05$ ).

L shell	Model									$R^2$	$r$
	DEMETER VLF	DEMETER VLF <sup>2</sup>	DEMETER VLF <sup>3</sup>	Illumination	IllumXVLF	Longitude	Latitude	Kp	KpXVLF		
L3	0.457*									0.209	0.457
	0.465*	0.118*	-0.058*							0.219	0.468
	.573*	0.212*	-0.026	-0.051*	0.033*	-0.099*	0.073*	0.027*	-0.254*	0.241	0.491
L4	0.444*									0.197	0.444
	0.328*	0.163*	0.058*							0.224	0.473
	0.374*	0.235*	0.089*	-0.049*	0.064*	-0.045*	0.079*	0.059*	-0.222*	0.240	0.490

649

650

651 Table 3. Nightside models with standardized regression coefficients predicting ground VLF (1.0 kHz) from satellite observations. Prediction  
 652 efficiency ( $R^2$ , fraction of variation in data explained by the regression model) and correlations ( $r$ ). \*: statistically significant coefficient ( $p < 0.05$ ).

L shell	Model									$R^2$	$r$
	DEMETER VLF	DEMETER VLF <sup>2</sup>	DEMETER VLF <sup>3</sup>	Illumination	IllumXVLF	Longitude	Latitude	Kp	KpXVLF		
L3	0.387*									0.149	0.386
	0.410*	0.091*	0.021							0.155	0.394
	0.569*	0.109*	0.009	-0.136*	0.024*	0.120*	0.018	-0.108*	-0.174*	0.198	0.445
L4	0.361*									0.130	0.361
	0.365*	0.098*	0.026							0.139	0.373
	0.527*	0.106*	0.040	-0.158*	-0.003	0.131*	0.050*	-0.084*	-0.197*	0.182	0.426

653

654

655

656

657

658

659 Table 4. Dawn models with unstandardized regression coefficients for calculation of predicted satellite observations from ground VLF (1.0 kHz).  
 660 Year 2006 was withheld as the test set. Validation correlations ( $r$ ) between VLF predicted by each model and observations from the test set are  
 661 given. Shrinkage is the difference between the coefficient of determination ( $R^2$ ) from the predictive model minus the square of the validation  
 662 correlation ( $r^2$ ) and quantifies the difference in variation explained by the model in the training set vs. that in the test set.

663 \*: statistically significant coefficient ( $p < 0.05$ ).

L shell	Model										$R^2$ Coefficient of determination	$r$ Validation correlation	Shrinkage $R^2 - r^2$
	Intercept	Halley VLF	Halley VLF <sup>2</sup>	Halley VLF <sup>3</sup>	Illumination	Illum×VLF	Kp	Kp×VLF	Longitudinal Distance	Hemisphere			
L3	-0.3759*	0.6129*									0.369	0.603	0.005
	-1.851*	3.178*	-0.9207*	0.08470*							0.526	0.709	0.023
	-1.877*	3.221*	-0.9319*	0.08562*	0.0008651	-0.001349					0.528	0.707	0.028
	-1.890*	2.699*	-0.8040*	0.07504*	-0.002478	-0.002437	0.02926*	0.0006072			0.598	0.764	0.014
	-1.923*	2.735*	-0.8296*	0.07870*	-0.001059	-0.003355*	0.02775*	0.0010580	-0.00000541	-0.07420*	0.589	0.765	0.004
L4	-0.167	0.5403*									0.318	0.517	0.051
	-1.416	2.731*	-0.8084*	0.07764*							0.430	0.630	0.033
	-1.356	2.738*	-0.8126*	0.07802*	-0.004007	-0.0005874					0.433	0.628	0.039
	-1.350	2.363*	-0.7300*	0.07119*	-0.006792*	-0.001669	0.01959*	0.001696			0.477	0.657	0.045
	-1.301*	2.371*	-0.7340*	0.07158*	-0.006822*	-0.001632	0.01934*	0.001782	-0.0002462	-0.05365	0.479	0.656	

664

665 Table 5. Dayside models with unstandardized regression coefficients for calculation of predicted satellite observations from ground VLF (1.0  
 666 kHz). Year 2006 was withheld as the test set. Validation correlations ( $r$ ) between VLF predicted by each model and observations from the test  
 667 set are given, along with the shrinkage: the difference in variation explained by the model in the training set vs. that in the test set.

668 \*: statistically significant coefficient ( $p < 0.05$ ).

L shell	Model										R <sup>2</sup> Coefficient of determination	r Validation corr
	Intercept	Halley VLF	Halley VLF <sup>2</sup>	Halley VLF <sup>3</sup>	Illumination	Illum×VLF	Kp	Kp×VLF	Longitudinal Distance	Hemisphere		
L3	-0.1806	0.4951*									0.203	0.468
	-1.313*	2.511*	-0.7400*	0.06854*							0.298	0.545
	-1.243*	2.480*	-0.7376*	0.06830*	-0.00634*	0.002404*					0.300	0.541
	-1.487*	1.849*	-0.5641*	0.05310*	-0.00924*	0.002031*	0.039670*	0.001312			0.485	0.682
	-1.543*	1.780*	-0.5458*	0.05188*	-0.00906*	0.001933*	0.004052*	0.001110	0.0007454*	-0.1081*	0.481	0.687
L4	-0.05233*	0.4690*									0.193	0.451
	-0.9906*	2.344*	-0.7031*	0.06711*							0.275	0.532
	-0.8872*	2.293*	-0.6947*	0.06160*	-0.00902*	0.003128*					0.279	0.528
	-1.097*	1.766*	-0.5475*	0.05314*	-0.01223*	0.003175*	0.03590*	0.0008672			0.422	0.628
	-1.002*	1.783*	-0.5569*	0.05420*	-0.01226*	0.002865*	0.03309*	0.0011530	-0.001162*	-0.03135*	0.426	0.631

669

670

671

672 Table 6. Nightside models with unstandardized regression coefficients for calculation of predicted satellite observations from ground VLF (1.0  
 673 kHz). Year 2006 was withheld as the test set. Validation correlations ( $r$ ) between VLF predicted by each model and observations from the test  
 674 set are given, along with the shrinkage: the difference in variation explained by the model in the training set vs. that in the test set.

675 \*: statistically significant coefficient ( $p < 0.05$ ).

L shell	Model										$R^2$ Coefficient of determination	$r$ Validation correlation	Shrinkage $R^2 - r^2$
	Intercept	Halley VLF	Halley VLF <sup>2</sup>	Halley VLF <sup>3</sup>	Illumination	Illum×VLF	Kp	Kp×VLF	Longitudinal Distance	Hemisphere			
L3	-0.8326*	0.4391*									0.140	0.405	-0.024
	-1.641*	1.982*	-0.6194*	0.06082*							0.198	0.476	-0.029
	-1.784*	2.154*	-0.6937*	0.06915*	0.01372*	0.002731					0.208	0.490	-0.032
	-2.034*	2.186*	-0.6692*	0.0659*	0.01090*	0.003533	0.01405*	-0.004226*			0.230	0.506	-0.026
	-1.991*	2.157*	-0.6463*	0.06327*	0.01714*	0.001876	0.01489*	-0.005069	-0.0009454*	-0.002568*	0.245	0.512	-0.017
L4	-0.6270*	0.4003*									0.119	0.381	-0.026
	-1.250*	1.587*	-0.4685*	0.04500*							0.155	0.427	-0.027
	-1.444*	1.806*	-0.5565*	0.05449*	0.001982*	0.0002783					0.169	0.457	-0.040
	-1.675*	1.893*	-0.5412*	0.05207*	0.01741*	0.001554	0.001263*	-0.006152*			0.182	0.465	-0.034
	-1.506*	1.909*	-0.5507*	-0.05336*	0.01755*	0.001029	0.01234*	-0.005898*	-0.001605*	-0.1600*	0.200	0.479	-0.029

676 Figure 1. Correlations of Halley (0.5, 1.0, 2.0, and 4.25 kHz channels) with DEMETER VLF. White: dayside  
677 (0600-1800 LT Halley; 1030 LT DEMETER), light gray: dawn (0600-0900 MLT Halley; 1030 LT DEMETER);  
678 Dark gray: nightside (1800-0600 LT Halley; 2230 LT DEMETER).

679  
680 Figure 2. Regression of Halley VELOX (centered at 1 kHz; 0.5-1.5 kHz) predicted by DEMETER VLF (0.5-  
681 1.5kHz). a. L3 - Dayside (Halley 0600-1800 LT, DEMETER 1030 LT); b. L3 -Dawn (0600-0900 MLT,  
682 DEMETER 1030 LT); c. L3 - Nightside (Halley 1800-0600 LT, DEMETER 2230 LT); d. L4 - Dayside; e. L4 -  
683 Dawn; f. L4 - Nightside.

684  
685 Figure 3. Standardized regression coefficients for models predicting ground Halley VLF activity (1 kHz)  
686 from DEMETER and Illumination (solar degrees above the horizon at Halley noon),  
687 IlluminationXDEMETER VLF interaction, hemisphere where DEMETER measurements are taken (South =  
688 1, North = 0), longitudinal separation between Halley and DEMETER, Kp, and the KpXDEMETER  
689 interaction. a. Dawn (Halley 0600-0900 MLT; 1kHz), DEMETER L3; b. Dayside (Halley 0600-1800 LT); c.  
690 Nightside (Halley 1800-0600 LT).  
691 \*: coefficient is statistically significant ( $p < 0.05$ ).

692  
693 Figure 4. Graphical representation of the interaction terms. a. Linear relationship between satellite  
694 (DEMETER, L3) and dawn ground (Halley, 0600-0900 MLT) VLF when sun elevation is less than the  
695 median (<5.9 degrees above horizon, gray line) and greater than the median (>5.9 degrees above  
696 horizon, black line). b. Relationship between DEMETER (L3) and dayside ground (Halley, 0600-1800 LT)  
697 VLF during od when Kp <2.3 (gray) and >2.3 (black).

698  
699 Figure 5. Standardized regression coefficients for models predicting ground Halley VLF activity 4.25 kHz)  
700 from DEMETER and Illumination (solar degrees above the horizon at Halley noon),  
701 IlluminationXDEMETER VLF interaction, hemisphere where DEMETER measurements are taken (South =  
702 1, North = 0), longitudinal separation between Halley and DEMETER, Kp, and the KpXDEMETER  
703 interaction. a. Dawn (Halley 0600-0900 MLT; 1kHz), DEMETER L3; b. Dayside (Halley 0600-1800 LT); c.  
704 Nightside(Halley 1800-0600 LT).  
705 \*: coefficient is statistically significant ( $p < 0.05$ ).

706  
707 Figure 6. Correlation of DEMETER satellite observations (L3; 0.5-1.5 kHz) with activity predicted by Halley  
708 dawn chorus (0600-0900 MLT) and various other parameters: a. Halley VELOX data linear model, b.  
709 Halley cubic model; c. Halley cubic model with solar illumination, Kp, and the illuminationXVLF and  
710 KpXVLF interaction terms. Year 2006 is held out as the test set while the remaining data is used to  
711 produce the model.

Figure 1.

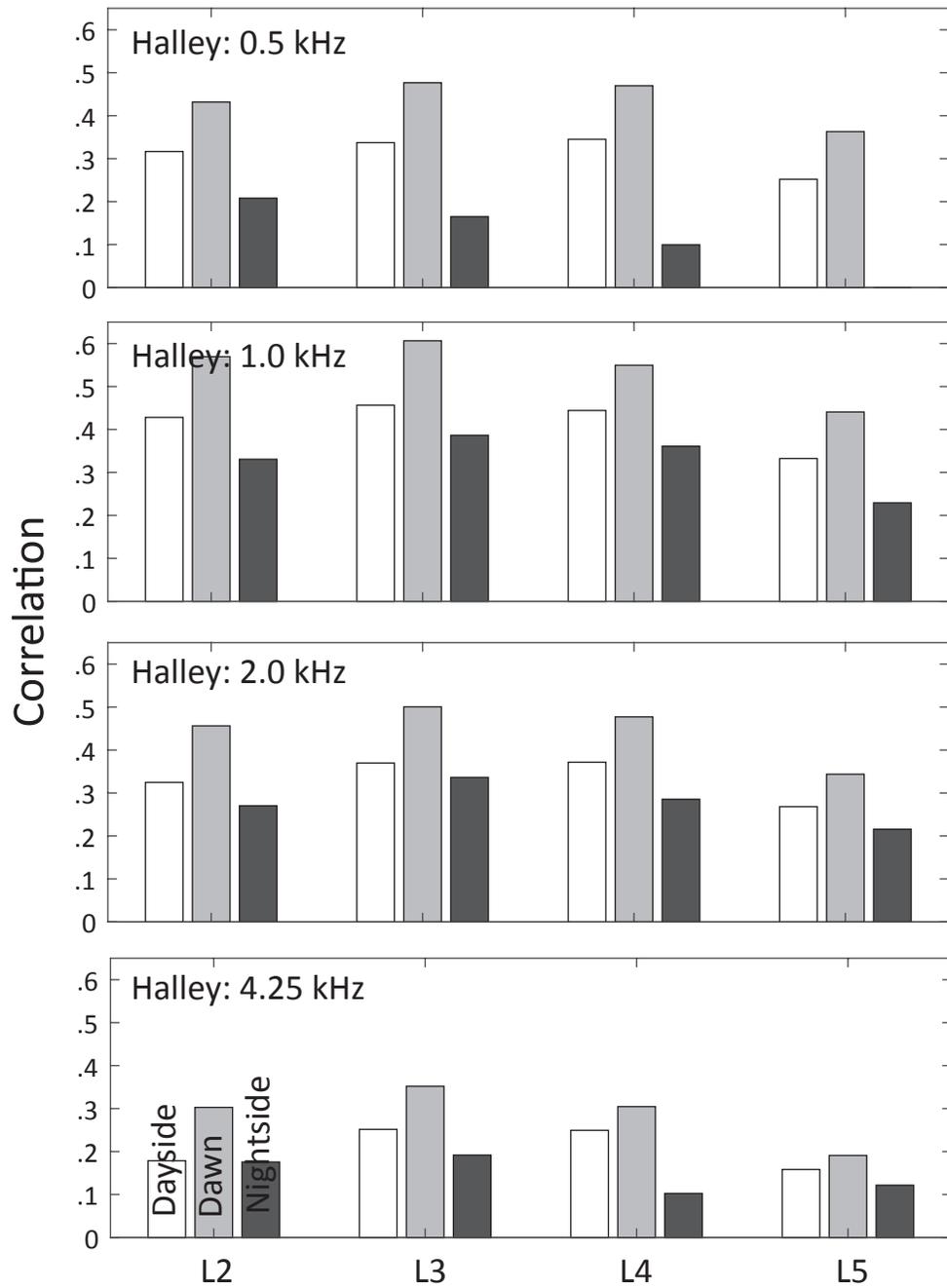


Figure 2.

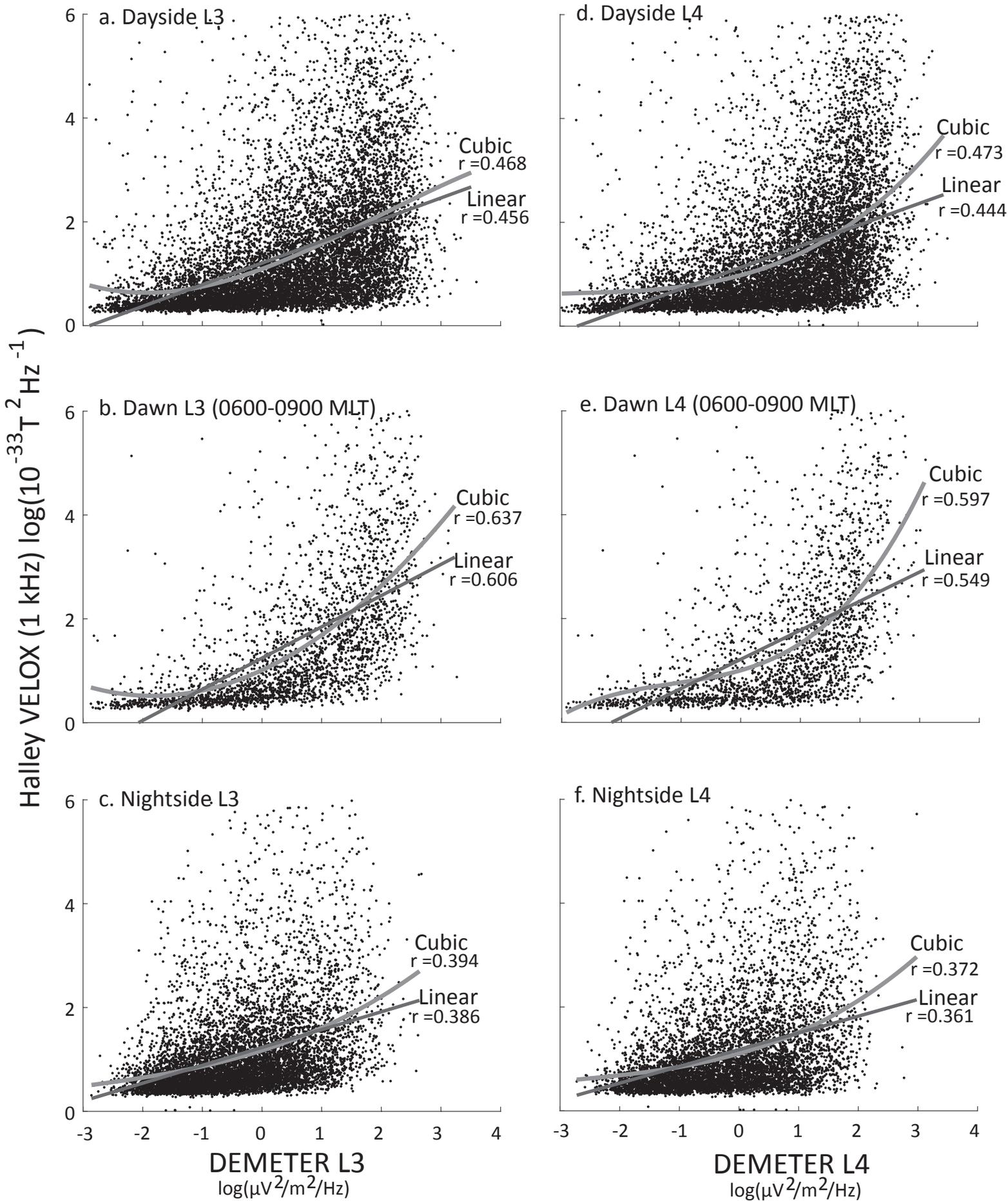
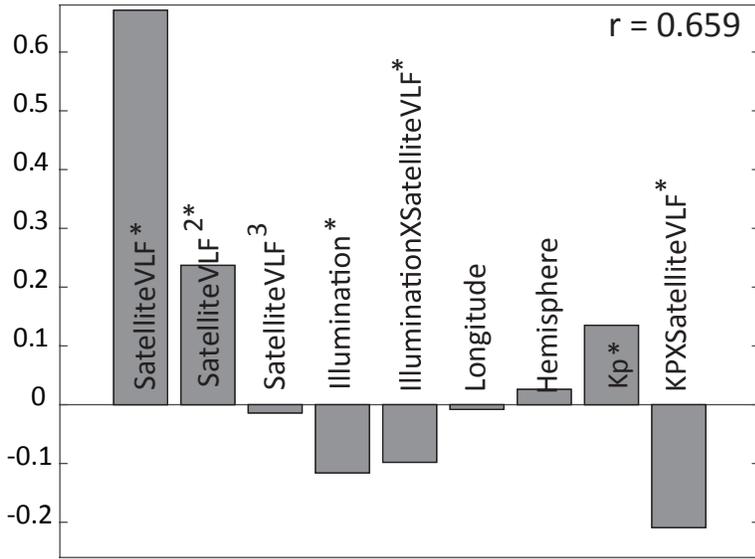


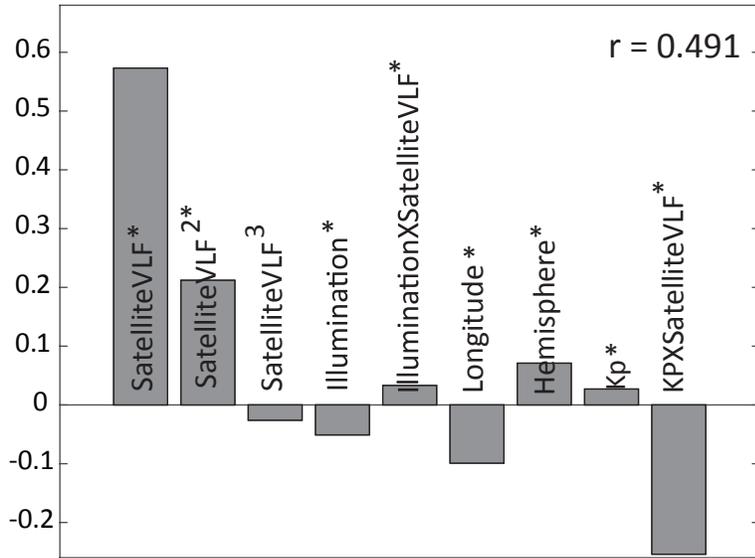
Figure 3.

Standardized Regression Coefficients

a. Dawn L3



b. Dayside L3



c. Nightside L3

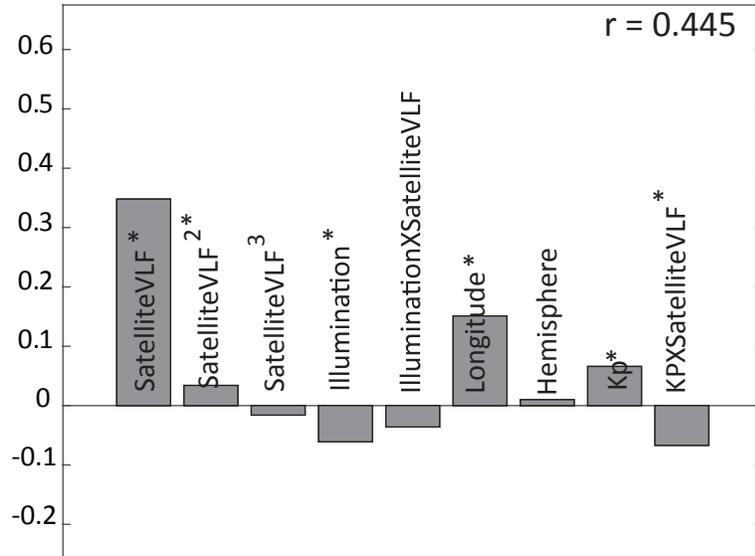


Figure 4.

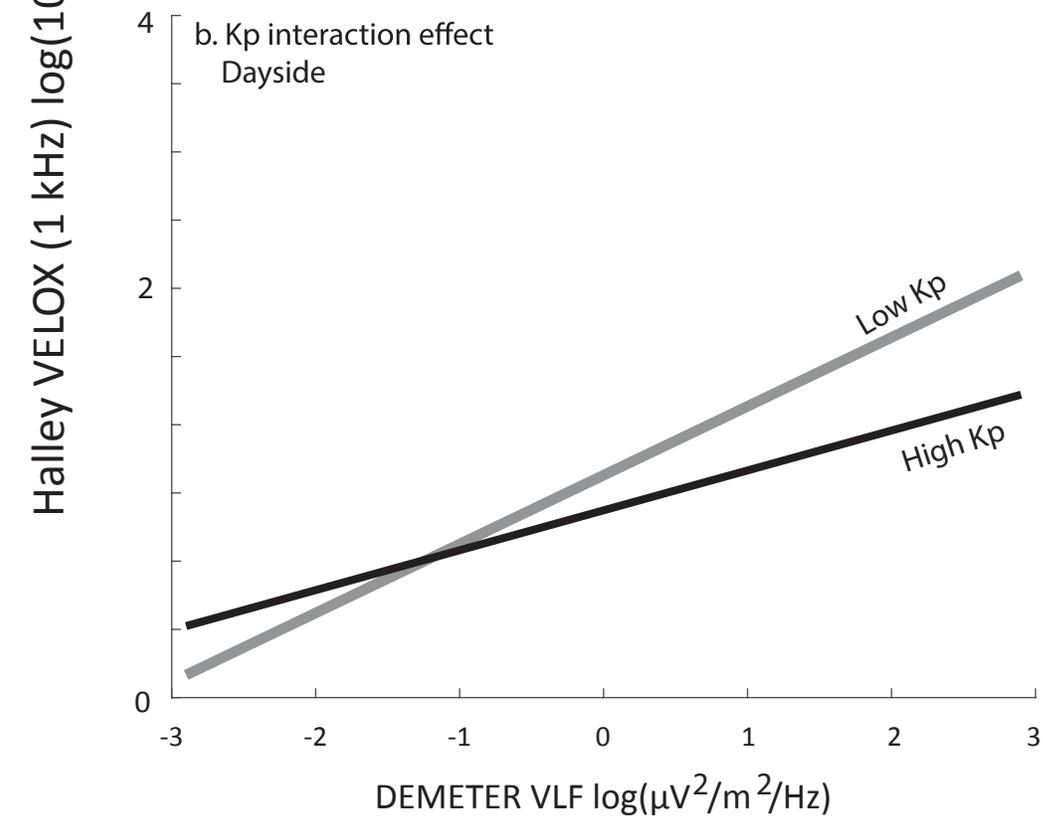
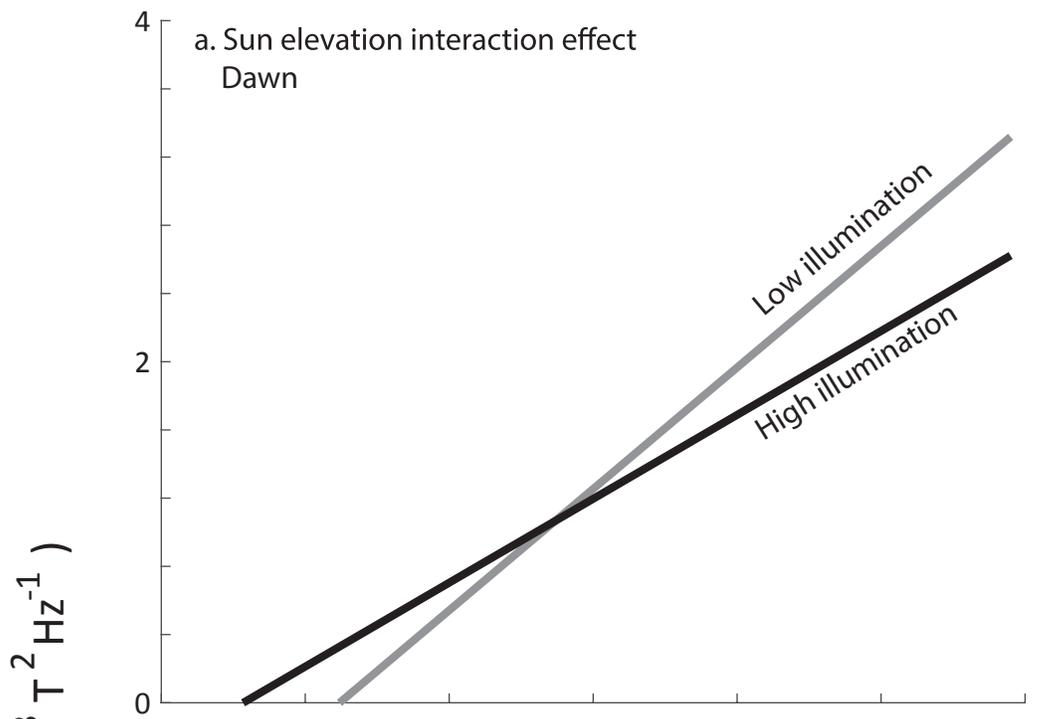
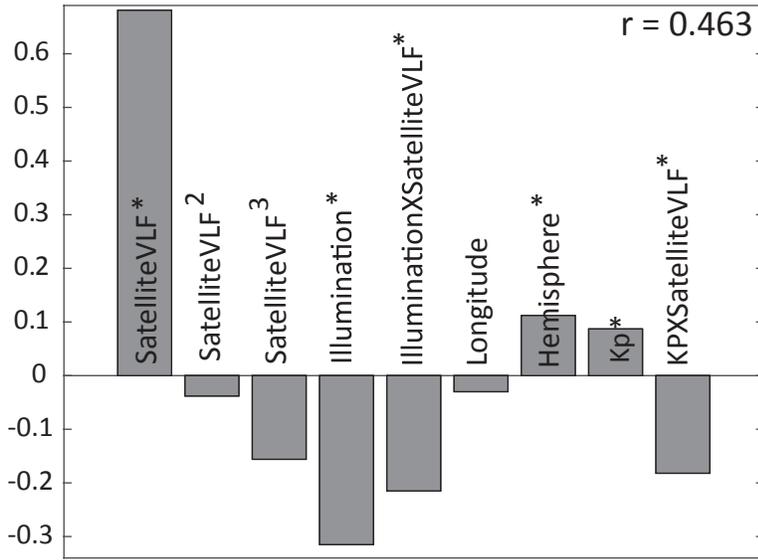


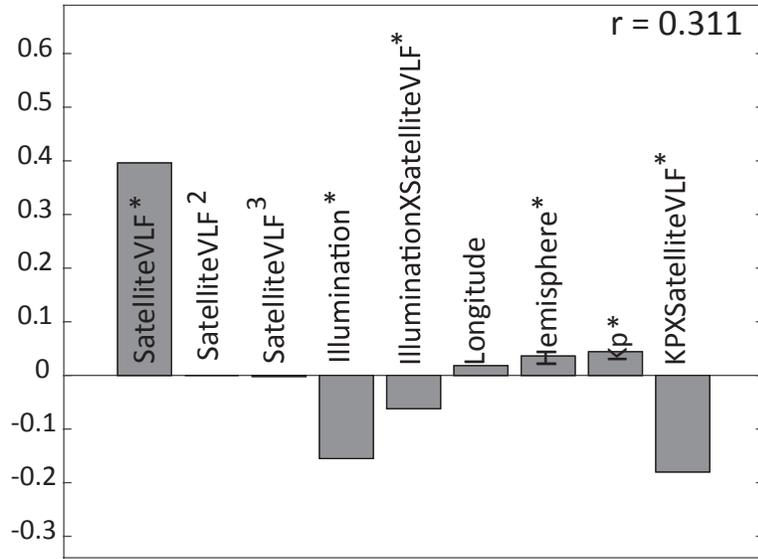
Figure 5.

Standardized Regression Coefficients

a. Dawn L3



b. Dayside L3



c. Nightside L3

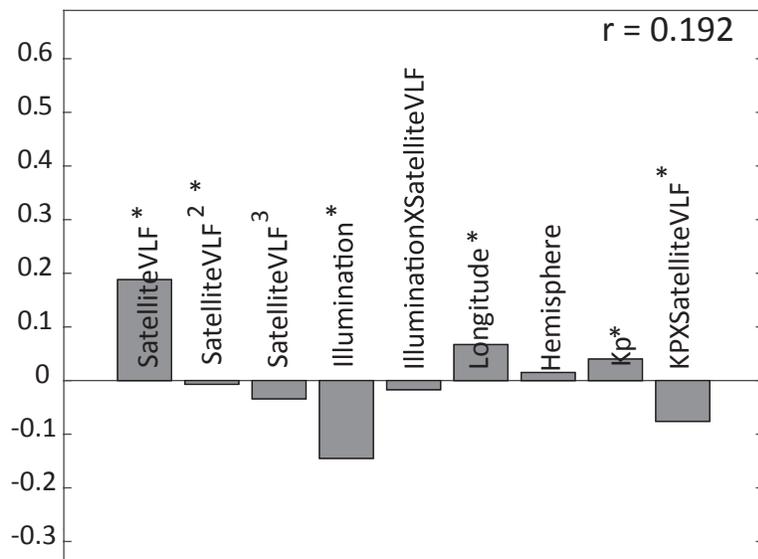
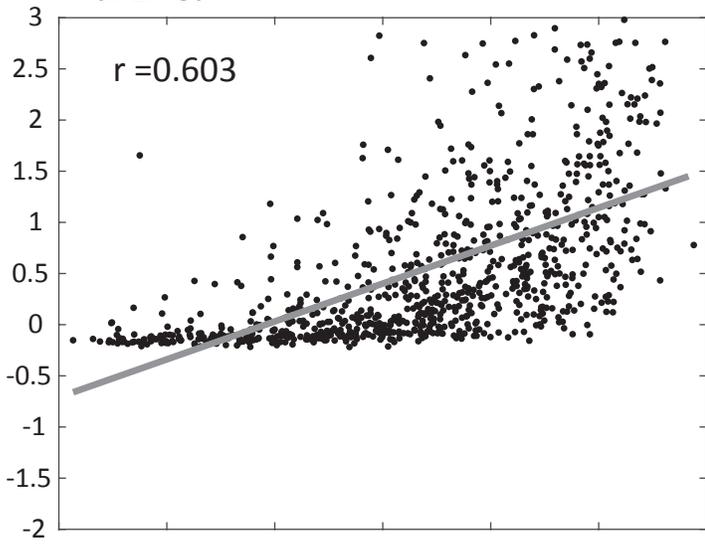
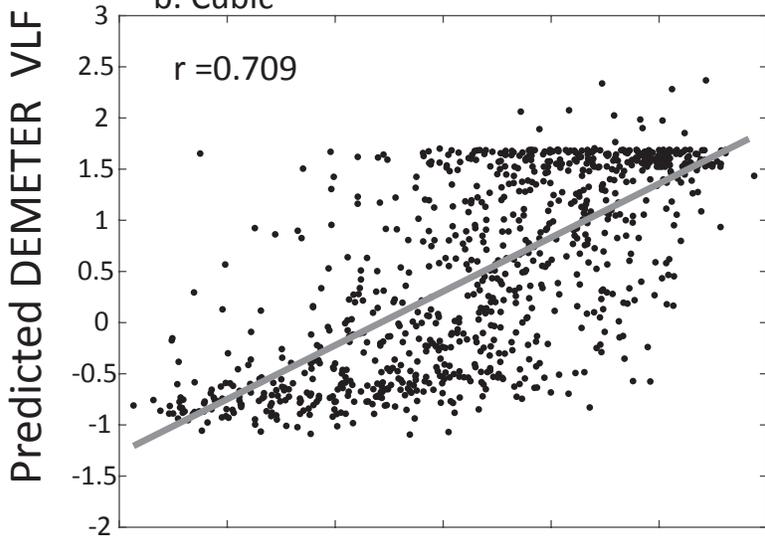


Figure 6.

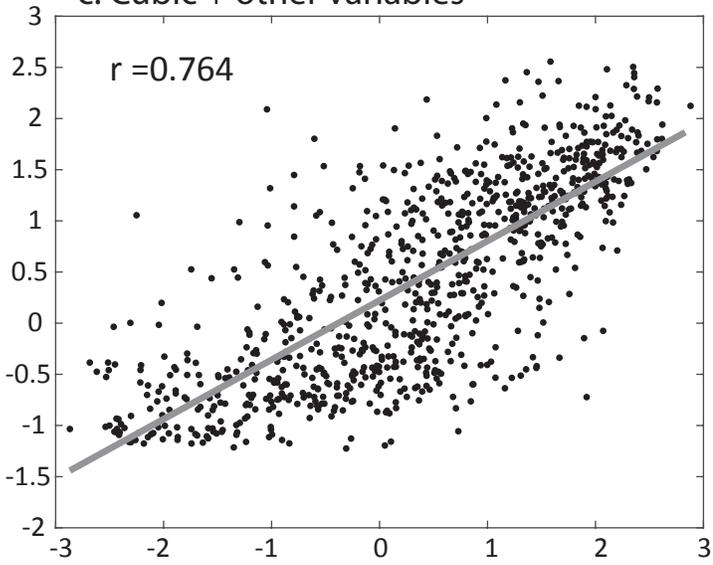
a. Linear



b. Cubic



c. Cubic + other variables



Observed DEMETER VLF