Supporting Information

Dynamical evidence for causality between galactic cosmic rays and global temperature

Anastasios A. Tsonis[1*], Ethan R. Deyle[2], Robert M. May[3], George Sugihara[2*], Kyle L. Swanson[1], Joshua D. Verbeten[1], Geli Wang[4]

1. Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA
2. Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA
3. Department of Zoology, University of Oxford, Oxford UK OX1-3PS
4. LAGEO Institute for Atmospheric Physics, Chinese Academy of Sciences, Beijing, CHINA

* Corresponding authors

SI Contents:

1) **Videos Illustrating EDM, Takens Theorem and CCM Methods:**
   -Movie_S1: Introduction to Empirical Dynamics
   -Movie_S2: Takens Theorem
   -Movie_S3: Convergent Cross Mapping (CCM)
2) **SI Text**
   -Simplex projection
   -S-map
   -CCM
   -Granger Cauality

SI Text

A note on simplex projection and on determining embedding dimension

Simplex projection is a nearest-neighbor forecasting algorithm (1) that involves tracking the forward trajectory of nearby points in a lag coordinate embedding. An exploratory series of embedding dimensions (E) are used to evaluate the prediction, and the best E is chosen based on prediction skill (Figure S1). This embedding is then used in the S-map procedure.

Notes on the S-map test for nonlinear dynamics

To determine whether a time series reflects linear or nonlinear processes we compare the out-of-sample forecast skill of a linear model versus an equivalent nonlinear model. To do this, we apply a two-step procedure: 1) we use simplex-projection (1) to identify the best embedding dimension, and 2) we use this embedding in the S-map procedure (2) to

assess the nonlinearity of the time series.

S-maps are an extension of standard linear autoregressive models in which the coefficients depend on the location of the predictee $\boldsymbol{y}_t$ in an E-dimensional embedding. New coefficients are recalculated (from the library of a predictant set $\boldsymbol{X}$) by singular value decomposition (SVD) for each new prediction. In this calculation, the weight given to each vector in the library depends on how close that vector $\boldsymbol{x}_t$ is to the predictee $\boldsymbol{y}_t$. The extent of this weighting is determined by the parameter θ.

As above, we generate an E-dimensional embedding from points in the library using lagged coordinates to obtain an embedded time series with vectors $\boldsymbol{x}_t \in \mathbb{R}^{E+1}$, where $x_t(1) = 1$ is the constant term in the solution of Eq. (S2) below. Let the time series observation in the prediction set T_p time steps forward be $Y_{t+Tp}(1) = Y(t)$.

Then the forecast for Y_t is $\hat{Y}_t = \sum_{j=0}^{E} C_t(j) X_t(j)$ (S1)

For our analysis, we chose $T_P = 1$. For each E-dimensional predictee vector $\boldsymbol{y}_t$, $\boldsymbol{C}$ is solved by SVD using the library set as follows:

$$\boldsymbol{B} = \boldsymbol{AC}, \quad \text{(S2)}$$

where $B_i = w(\| \boldsymbol{x}_i - \boldsymbol{y}_t \|) y_i$, $A_{ij} = w(\| \boldsymbol{x}_i - \boldsymbol{y}_t \|) x_i(j)$, and $w(d) = e^{-\theta d_{it}/\bar{d}}$, $\theta \geq 0$, d_{it} is the distance between $\boldsymbol{y}_t$ and the i^{th} neighbor vector $\boldsymbol{x}_i$ in the library embedding, and the scale vector, $\bar{d}$, is the average distance between neighbors in the library. Note that $\boldsymbol{A}$ has dimension $n \times (E+1)$, where n = size of the library. Again, a different map is generated for each forecast, with the weightings in each map depending on the location of the predictee in the E-dimensional state space. This weighting procedure is governed by the tuning parameter θ, where $\theta = 0$ gives a global linear map, and

increasing values of θ give increasingly local or nonlinear mappings. When $\theta = 0$, all vectors are more or less weighted equally so a single (global) linear map can be used for all predictions. In the case where $\theta > 0$, vectors closest to the predictee in state-space are weighted more heavily in the SVD solution. Such forecasts emphasize local information in the library set, and are therefore nonlinear.

Notes of Convergent Cross Mapping (CCM)

Consider two time series of length L, $\{X\} = \{X(1), X(2), \ldots, X(L)\}$ and $\{Y\} = \{Y(1), Y(2),\ldots, Y(L)\}$. We begin by forming the lagged-coordinate vectors $\underline{x}(t) = \langle X(t), X(t\text{-}\tau), X(t\text{-}2\tau), \ldots, X(t\text{-}(E\text{-}1)\tau)\rangle$ for $t = 1+(E\text{-}1)\tau$ to $t = L$. This set of vectors is the "reconstructed manifold" or "shadow manifold" $\boldsymbol{M}_X$. To generate a cross-mapped estimate of $Y(t)$, denoted by $\hat{Y}(t) \mid \boldsymbol{M}_X$, we begin by locating the contemporaneous lagged-coordinate vector on $\boldsymbol{M}_X$, $\underline{x}(t)$, and find its E+1 nearest neighbors. Note that E+1 is the minimum number of points needed for a bounding simplex in an E-dimensional space. Next, denote the time indices (from closest to farthest) of the E+1 nearest neighbors of $\underline{x}(t)$ by t_1, $\ldots t_{E+1}$. These time indices corresponding to nearest neighbors to $\underline{x}(t)$ on $\boldsymbol{M}_X$ are used to identify points (neighbors) in Y (a putative neighborhood) to estimate $Y(t)$ from a locally weighted mean of the E+1 $Y(t_i)$ values.

$$\hat{Y}(t) \mid \boldsymbol{M}_X = \Sigma\, w_i\, Y(t_i) \qquad i = 1 \ldots E+1$$

where w_i is a weighting based on the distance between $x(t)$ and its ith nearest neighbor on $\boldsymbol{M}_X$ and $Y(t_i)$ are the contemporaneous values of Y. The weights are determined by

$$w_i = u_i / \Sigma\, u_j \qquad j=1 \ldots E+1$$

where

$$u_i = \exp\{-d[\underline{x}(t), \underline{x}(t_i)] / d[\underline{x}(t), \underline{x}(t_1)]\}$$

and $d[\underline{x}(s), \underline{x}(t)]$ is the Euclidean distance between two vectors. Cross mapping from Y to X is defined analogously.

Notes on Granger causality

According to Granger causality, given two simultaneously recorded time series X_t and Y_t where $t = 1, N$ denotes sampling times, we say that Y has causal influence on X if the variance of the prediction error of X given Y is less than the variance of the prediction of X not given Y**.** This means that if prediction of some output improves with the addition of an input, then the input Granger-causes the output. In its original formulation Granger causality is based on linear prediction of stochastic time series.

There are several ways to test for Granger causality. The approach used here uses the autoregressive specification of a bivariate vector autoregression. For a given lag m, we estimate the following unrestricted equation by ordinary least squares:

$$X_t = c + \sum_{i=1}^{m} a_i X_{t-i} + \sum_{i=1}^{m} b_i Y_{t-i} + e_t$$

Where a, b, and c are coefficients and e is a residual. The null hypothesis that '*Y does not Granger-cause X*' is then constructed as

$$H_o: b_1 = b_2 = \ldots = b_m = 0$$

We also estimate the equation

$$X_t = c + \sum_{i=1}^{m} a_i X_{t-i} + w_t$$

and compare the sum of squared residuals

$$RSS_1 = \sum_{t=1}^{N} \hat{e}_t^2$$

and

$$RSS_2 = \sum_{t=1}^{N} \hat{w}_t^2$$

The statistic $S = \frac{(RSS2-RSS1)}{RSS1} \frac{(T-2m-1)}{m}$ is approximately equal to $F_{m,T-2m-1}$, and it is statistically significant at a p level of

$$p = 1 - prob(F_{m,T-2m-1})$$

In our case if X is ΔGT and Y is CR the p value assuming an AR-1 (AR-2) process is 0.82 (0.97). If X is CR and Y is ΔGT, the respective p values are 0.81 and 0.64. Thus neither X nor Y Granger-causes the other. In fact, the variance explained in the prediction error is less than 1% regardless which variable is used to predict the other. Considering higher order AR processes does not improve these results.

References

1. Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* 344: 733-741.
2. Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. *Phil. Trans. R. Soc. London Ser. A* 348: 477-495.

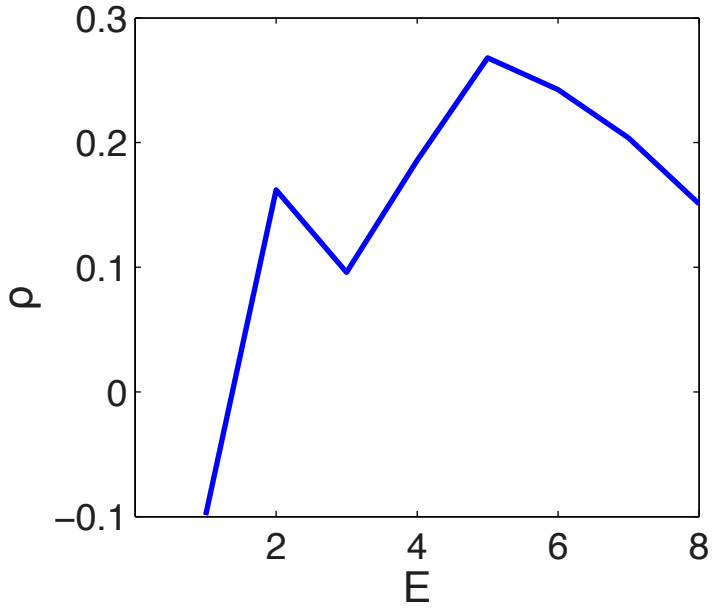

Figure S1: *Prediction skill (correlation coefficient between actual and predicted values) as a function of the embedding dimension (E) for the first-differenced GT time series using the nonlinear prediction method of Sugihara and May (1). The results indicate nonlinear dynamics with an optimum embedding dimension of 5.*

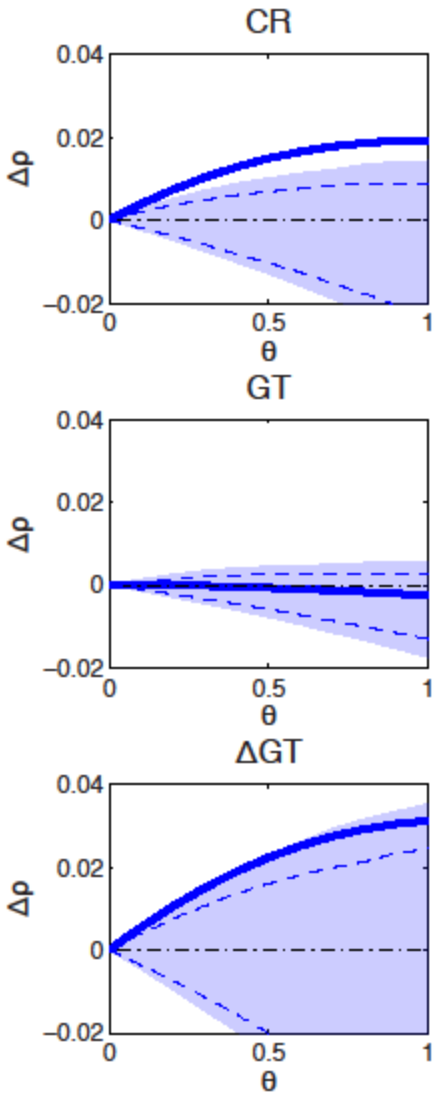

Figure S2. ***(Fig 2 in main text).*** *The S-map analysis of (a) the CR time series (aa proxy), (b) the GT time series, and (c) the first-differenced ΔGT time series. Δρ is the difference in the correlation between actual and predicted values between a linear model (global linear map) and an equivalent nonlinear model (local or nonlinear mappings). In a sense, it is a measure of the curvature of the manifold. Evidence for nonlinear dynamics is demonstrated if predictability improves as the S-map model parameter θ is tuned toward nonlinear solutions (θ >0). The shaded area is the 5^{th}/95^{th} and the dashed blue line the 10^{th}/90^{th} percentile confidence intervals using surrogate data (see text for details on surrogate data). The figure shows that while CR and ΔGT both show statistical nonlinear state dependent dynamics, the raw GT time series does not.*

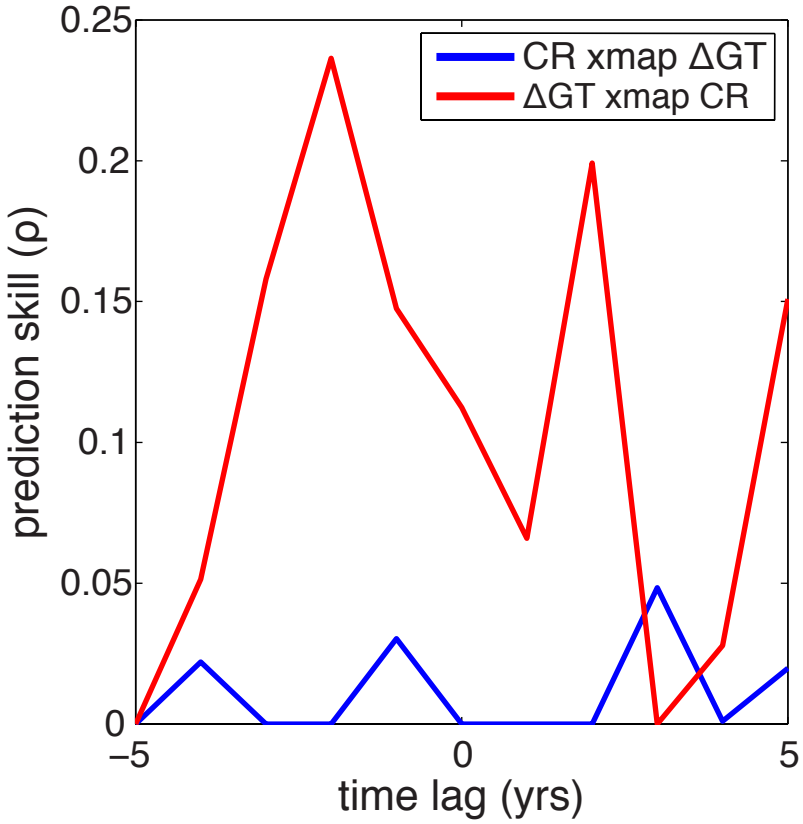

Figure S3: *CCM is repeated for different lags to determine the optimum lag. In the case of ΔGT cross map CR (red line) this lag is -2, which is sensible as the effect (of CR on ΔGT) cannot precede the cause. However in the meaningless case CR cross map ΔGT (non-significant blue line) this lag is 3.*

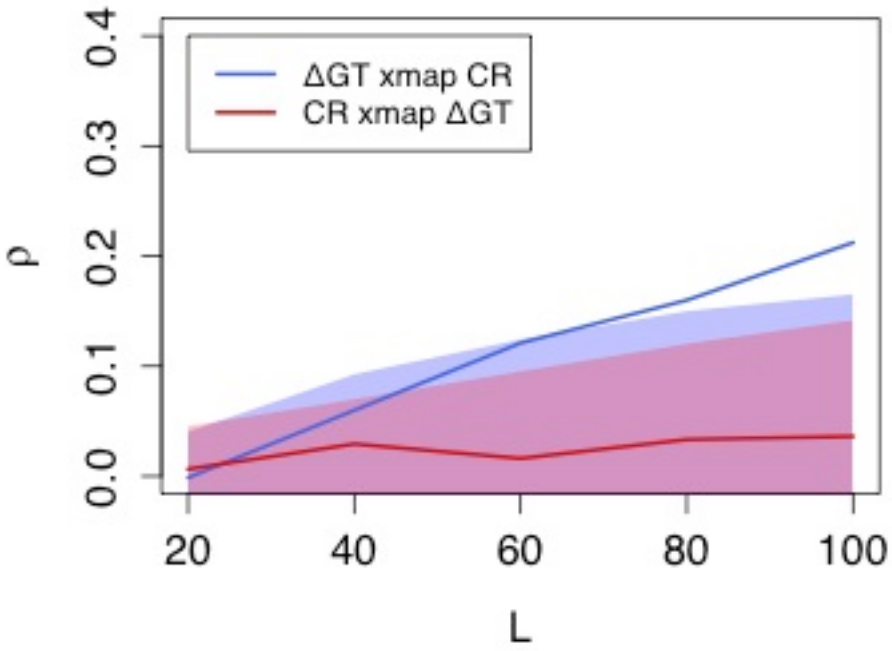

Figure S4: *As in Fig 3 of the main paper but with AR1 surrogates. Results of CCM analysis between the CR time series and the annual variations in global temperature (ΔGT). Convergence (increasing and significant ρ with longer time series) (blue line) shows that year-to-year changes in global temperature are causally forced by galactic cosmic rays. Lack of convergence (red line) shows, as expected, that ΔGT has no causal influence on CR. Results of the AR1 surrogates are nearly identical to those obtained by the Ebisuzaki method in the main text.*

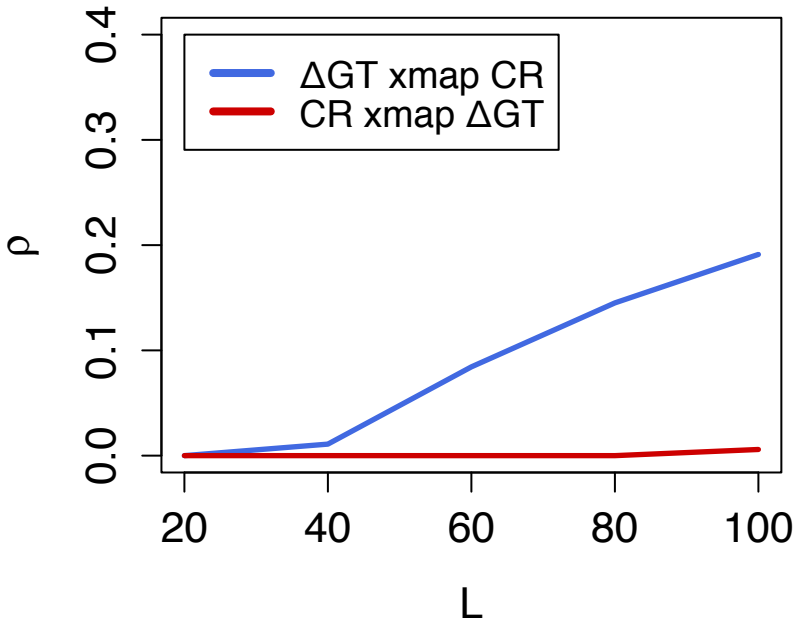

Figure S5: *Same as Fig. 3 in the main paper but for a suboptimal embedding dimension 8 to demonstrate robustness. Again we recover the results shown in Fig. 3.*

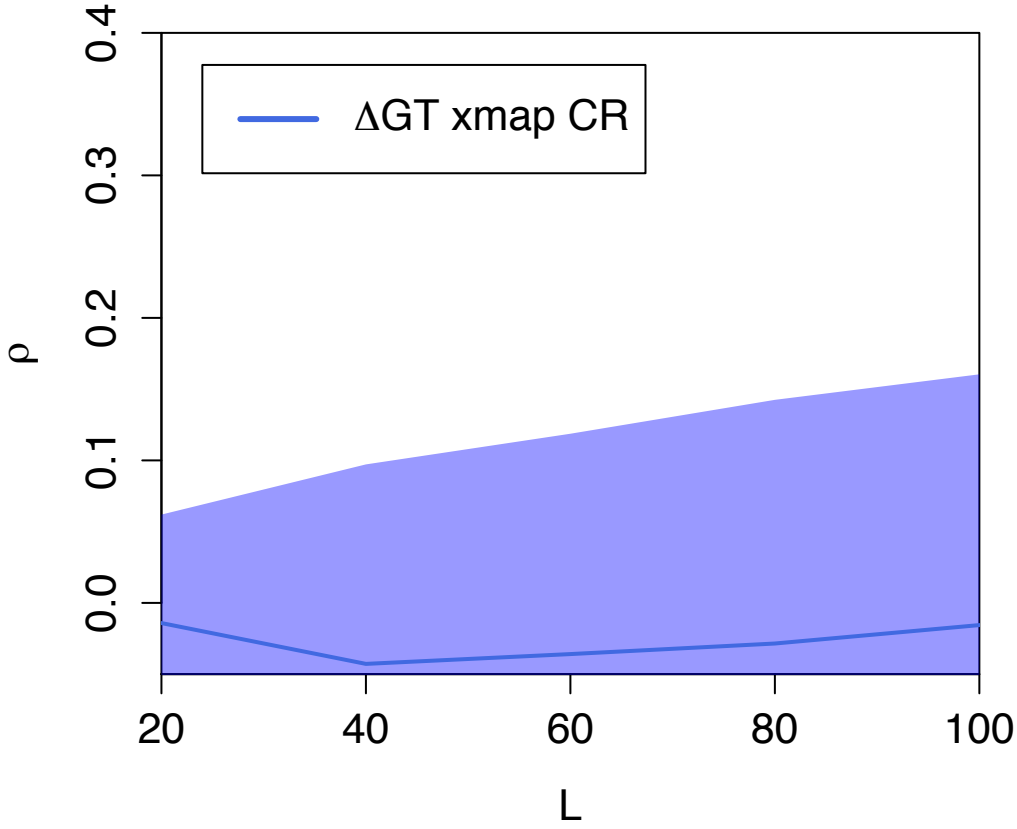

Figure S6: *Results of CCM analysis between the CR time series and the annual variations in global temperature (ΔGT) generated by the CCSM4 NCAR model (an IPCC AR5 model). As expected, because this model does not include any mechanism for cosmic rays to affect temperature, CCM shows there is no identifiable causality—there is no significant cross-mapping between the historical cosmic rays time series using ΔGT from the model. Furthermore, the result falls in the middle of the null expectation intervals based on AR1 surrogates (as used in Fig. S4 for the historical temperature data).*

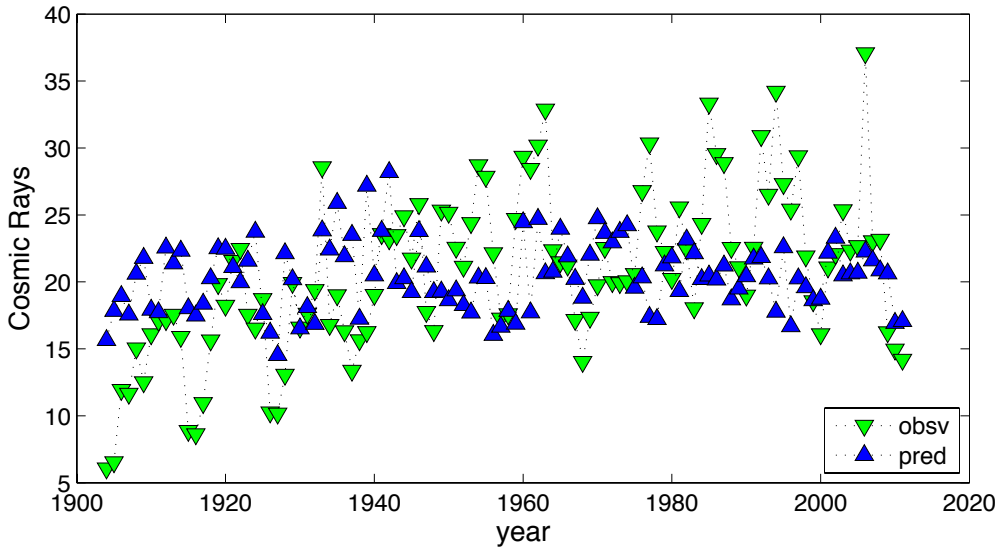

Figure S7: *Cross mapping results shown as observed vs predicted along the CR time series. The overall correlation coefficient is 0.20 in accordance with the blue line in Fig. S4.*

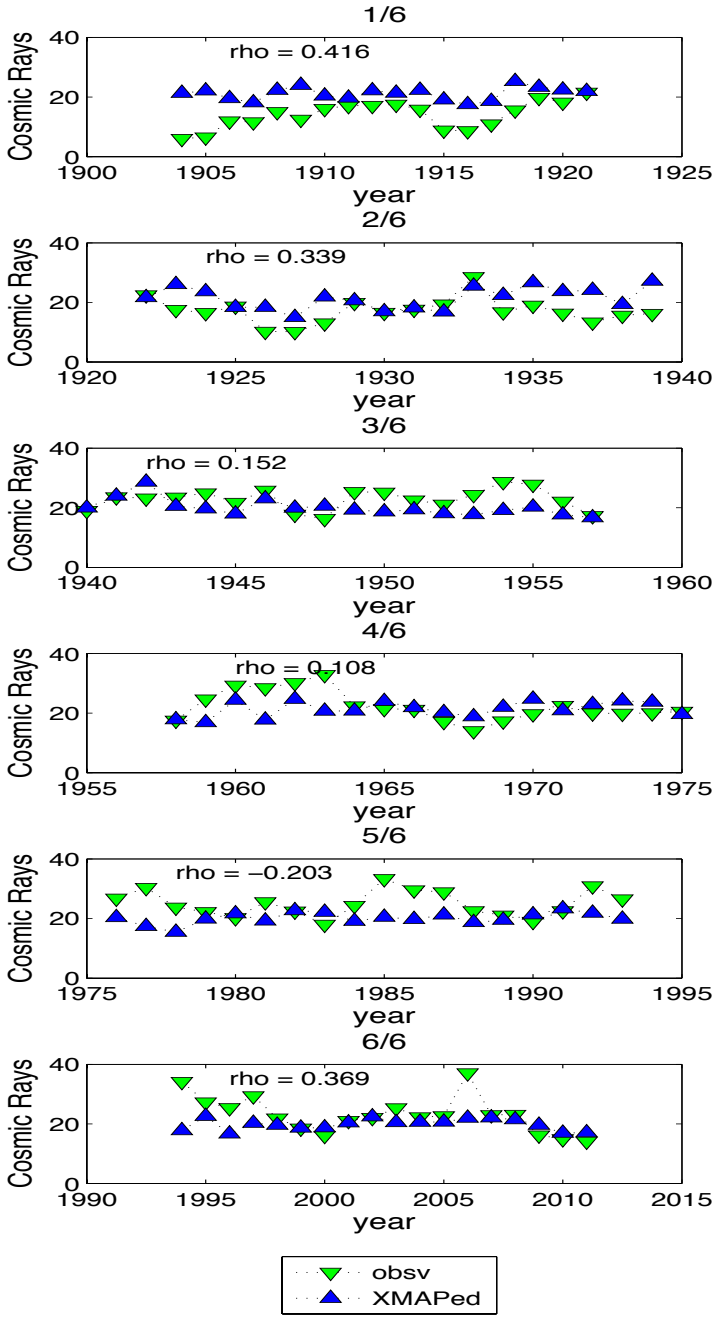

Figure S8: *Cross mapping done with 6-fold cross validation, were the cross map period is held out of sample from the libraries used to predict the out of sample period. We observe that in five out of six sub-periods the correlation coefficient between actual and predicted values is positive, as one would expect when CR causes ΔGT. Given the small*

sample sizes involved as well as the fact that the effect of CR on ΔGT is weak to moderate, this result provides additional confidence about the causality between CR and ΔGT.