SUPPORTING INFORMATION

Supporting Information for: Macrophenological dynamics from citizen science plant occurrence data

Karin Mora[1,2,3] | Michael Rzanny[4] | Jana Wäldchen[3, 4] | Hannes Feilhauer[1, 2, 3] | Teja Kattenborn[5] | Guido Kraemer[1, 2] | Patrick Mäder[3, 6, 7] | Daria Svidzinska[1, 2] | Sophie Wolf[1, 2] | Miguel D. Mahecha[1, 2, 3, 8]

[1]Institute for Earth System Science & Remote Sensing, Leipzig University, 04103 Leipzig, Germany

[2]Remote Sensing Centre for Earth System Research, Leipzig University, 04103 Leipzig, Germany

[3]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany

[4]Max Planck Institute for Biogeochemistry, 07745 Jena, Germany

[5]Department for Sensor-based Geoinformatics, University of Freiburg, 79106 Freiburg, Germany

[6]Data Intensive Systems and Visualisation, Technische Universität Ilmenau, 98693 Ilmenau, Germany

[7]Faculty of Biological Sciences, Friedrich Schiller University, 07743 Jena, Gemany

[8]Center for Scalable Data Analytics and Artificial Intelligence, ScaDS.AI Dresden/Leipzig, Germany.

S1 | DATA AND SCALING

Data points x are scaled by the maximum across all data points, $\max(x)$,

$$\text{Scaling}(x) = \frac{x}{\max(x)}. \tag{1}$$

This scaling has been applied to the number of total observations, the number of distinct locations, and the number of distinct species observed per time step t, respectively. The observed maximum values are shown in Table s1. The scaled data are shown in Fig. 2 (manuscript).
The total number of observations per time step t and not the scaled median observations per grid cell was chosen, as they have a linear relationship and hence reflect the same aspects of the spatial distribution (Fig. s1).

TABLE s1 Observation counts: minimum, maximum, and median values at time t.

Counts across Germany per time step t	Min	at t =	Max	at t =	Median
number of total observations	43,054	48	1,347,268	19	560,806
number of distinct locations	2684	48	2991	19	2976
number of distinct species	1622	50	2782	20	2552
max. no. of observations/species per cell	219	50	1503	17	907

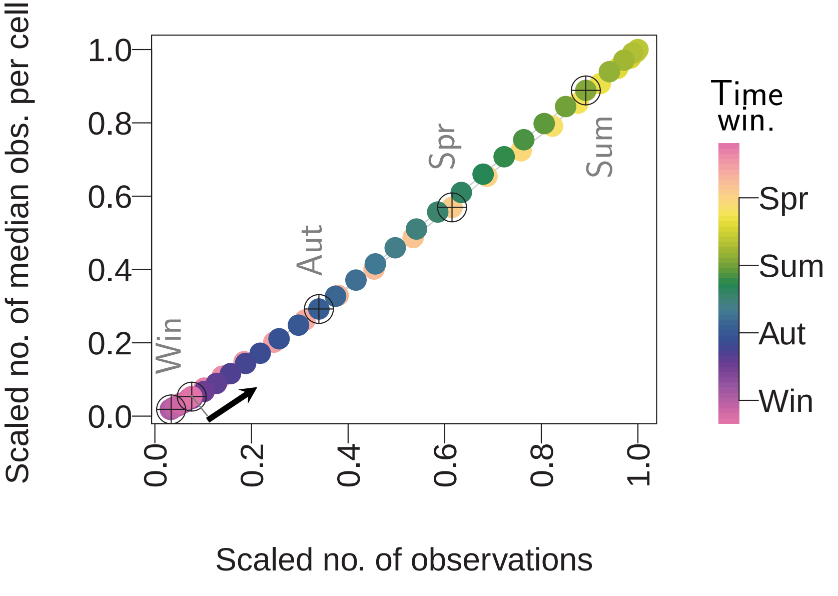

FIGURE s1 The total number of observations per time step t versus the scaled number of median observations per grid cell reflecting one aspect of spatial observation distribution. They have a linear relationship indicating their equivalence for our analysis.

S2 | MACROPHENOLOGICAL SERIES: ALTERNATIVE VISUALISATION

An alternative to the data cube visualisation of MPS1 to MPS4 (Fig. 3) is shown in Figs s2 – s5, where each MEP is shown.

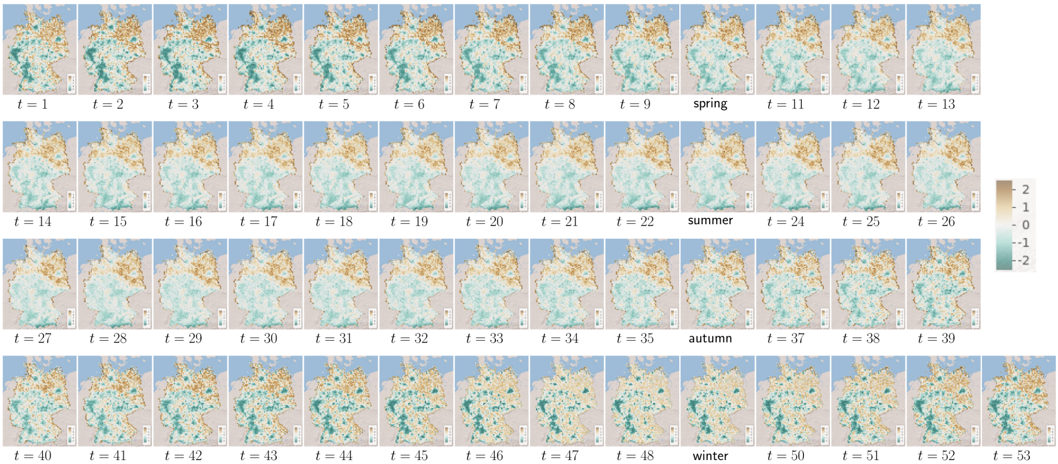

FIGURE s2 Alternative visualisation to Fig. 3 of MPS1 illustrating each MEP. The time step t indicates the starting week of each 90-day time window. Phase I observed during growing season: characteristic pattern representing environmental gradient. Phase II observed outside growing season: the MEPs are noisy and dominated by the human population distribution.

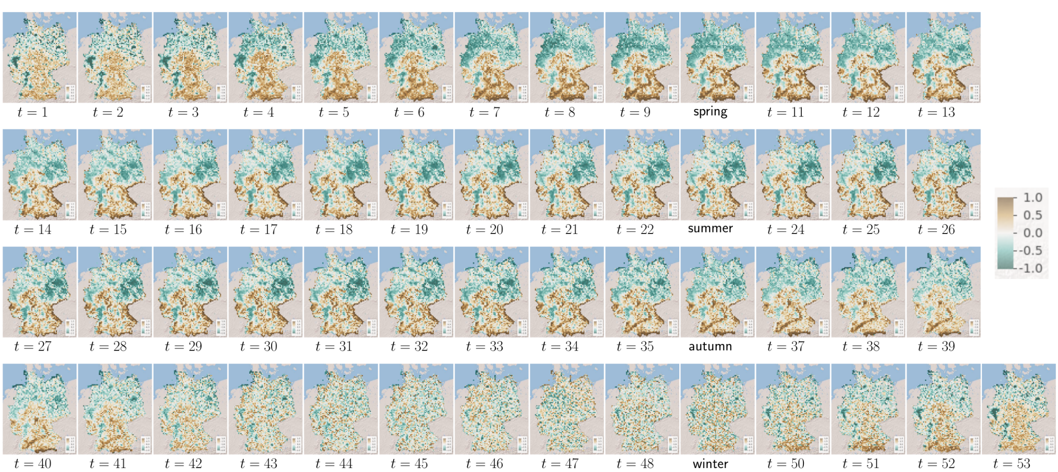

FIGURE s3 Alternative visualisation to Fig. 3 of MPS2 illustrating each MEP. The time step t indicates the starting week of each 90-day time window. Phase I observed during growing season: characteristic pattern representing environmental gradient. Phase II observed outside growing season: the MEPs deteriorate and become noisy with the human population distribution pattern being partly visible.

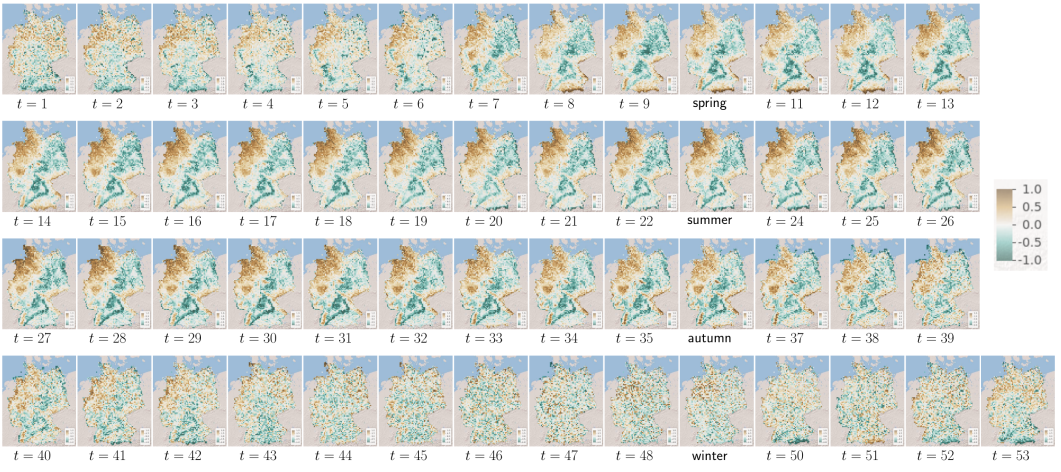

FIGURE s4 Alternative visualisation to Fig. 3 of MPS3 illustrating each MEP. The time step t indicates the starting week of each 90-day time window. Phase I observed during growing season: characteristic pattern representing environmental gradient. Phase II observed outside growing season: the MEPs deteriorate and become noisy

FIGURE s5 Alternative visualisation to Fig. 3 of MPS4 illustrating each MEP. The time step t indicates the starting week of each 90-day time window. Phase I observed during growing season: characteristic pattern representing environmental gradient. Phase II observed outside growing season: the MEPs deteriorate and become noisy

S3 | RESIDUAL VARIANCE SCALING

The residual variance corresponding to dimension one is shown in Table s2. These are the scaling values for Fig. 4.

TABLE s2 The residual variance (Res. Var) corresponding to dimension one per time step t. Note, that the static FI data has a Res. Var. corresponding to dimension one of 0.4954. This is used to scale the respective residual variance values in Fig. 4.

t	**Res. Var.**	t	**Res. Var.**	t	**Res. Var.**	t	**Res. Var.**
1	0.3825	16	0.5149	31	0.6023	46	0.5882
2	0.3166	17	0.5112	32	0.6148	47	0.6309
3	0.3125	18	0.5137	33	0.6224	48	0.6351
4	0.3096	19	0.5111	34	0.6336	49	0.5797
5	0.3206	20	0.5068	35	0.6437	50	0.5438
6	0.3406	21	0.5186	36	0.6365	51	0.5075
7	0.3703	22	0.5161	37	0.6417	52	0.4544
8	0.397	23	0.5286	38	0.6263	53	0.3945
9	0.4306	24	0.53	39	0.6063		
10	0.4629	25	0.5265	40	0.5972		
11	0.4858	26	0.539	41	0.557		
12	0.5091	27	0.5533	42	0.5288		
13	0.513	28	0.5642	43	0.5108		
14	0.5095	29	0.5891	44	0.5293		
15	0.5192	30	0.6032	45	0.5509		

S4 | SCALED RESIDUAL VARIANCE VS. COUNTS

The relationship of observation counts and scaled residual variance across Germany per time step is robust across Isomap dimensions, Fig. s6.

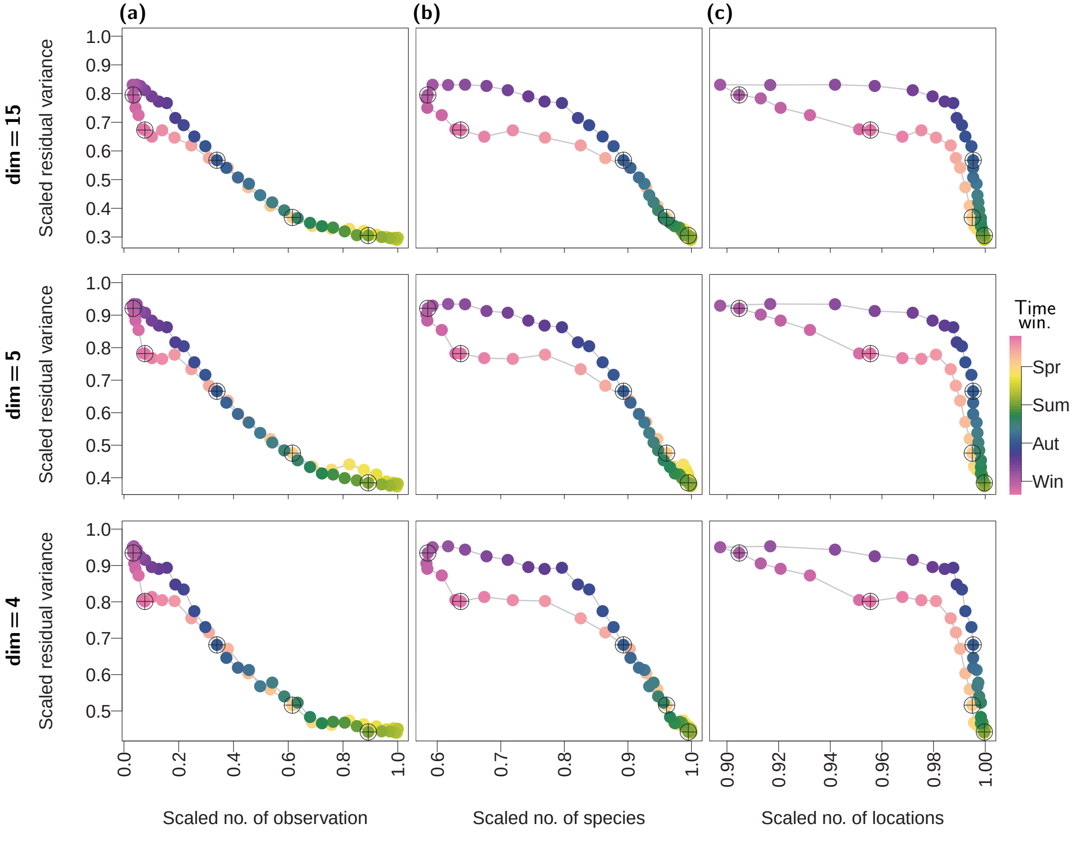

FIGURE s6 Nonlinear relationship between the scaled residual variance for embedding dimensions $p = 4, 5, 15$ and (a) scaled number of observations, (b) scaled number of locations, and (c) scaled number of species. Low scaled variance indicates high data compressibility, high scaled residual variance indicates low data compressibility. Crosses indicate the time windows corresponding to the meteorological seasons.
Even if the counts are the same, seasonal effects can be observed in the scaled residual variance, specifically pre- and post-winter (dark blue, magenta). The changes pre- and post-summer are minor in comparison (yellow, green). The nonlinear effects are consistently observed across embedding dimensions.

S5 | ISOMAP COMPUTATIONS OF AGGREGATED TIME SERIES WITH A WINDOW LENGTH OF 45 DAYS

In this section, we analyse a time series of occurrence data, which has been aggregated with a time window size (TWS) of 45 days, instead of 90 days as in the main analysis. Note that time windows have the same starting date but different end dates, see Table s3 for a selection of the time windows. We compare the scaled residual variance Fig. s7 and macrophenological series MPS2 s8. The main consequence of reducing the TWS, can be observed in spring. Then the residual variance remains high until mid-spring. This indicates that data compressibility remains low, which indicates that synchronised plant group behaviour emerges later and changes fast once it emerges. In contrast, during the autumn period the TWS has little effect on the scaled residual variance and data compressiblity. This indicates that the changes in group behaviour are equally well captured by a TWS of 45 or 90 days. This is also reflected in the macroecological patterns in Fig. s8.

TABLE s3 A selection of the time windows resulting from applying a time window size (TWS) of 45 and 90 days, respectively. Date format is day/month

time step t	TWS = 45	TWS = 90
1	1/1-14/2	1/1-30/3
10	4/3-17/4	4/3-1/6
23	3/6-17/7	3/6-31/8
36	2/9-16/10	2/9-30/11
49	2/12-15/1	2/12-1/3

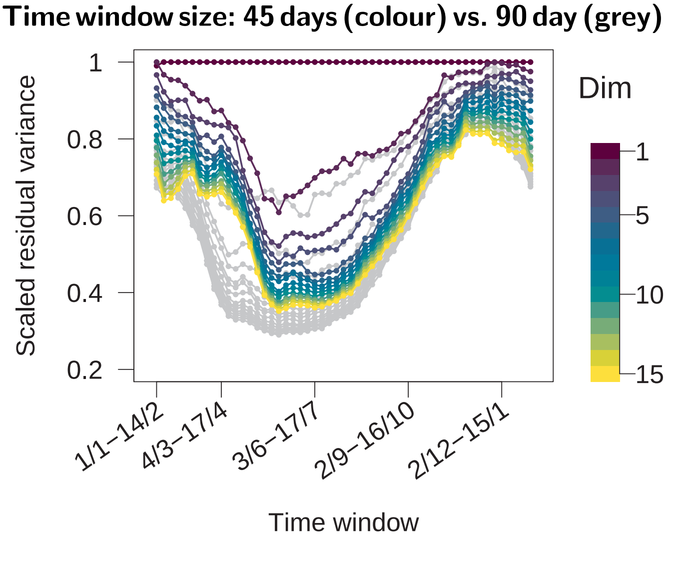

FIGURE s7 Scaled residual variance (sRV): comparison for the spatio-temporal analysis with time window size (TWS) 45 days and 90 days. The data compressiblity in spring is worse (higher sRV) with TWS = 45 than with TWS = 90. This is a result of plants occurring later towards mid spring. The autumn period is comparable to between TWS=45 and TWS=90. This indicates that the spatio-temporal species occurrence and hence the synchronised group behaviour of plants is similar irrespective of the TWS.

FIGURE s8 Macrophenological series (MPS2) comprised of macroecological pattern (MEP2) per time step t. Outside the growing season the noisy MEPs last into early spring: $t = 6$ corresp. to [5/2-20/3] for TWS=45 compared to TWS=90 with [5/2-4/5]. The pattern deterioration in MPS2 computed with TWS=45 occurs over a similar time period, i.e. t=40 corresp. to [30/9-13/11] with TWS=45 and [30/9-28/12] with TWS=90. This indicates that the synchronised group behaviour of plants is similar in autumn irrespective of the TWS.

S6 | CANONICAL VARIATES OF MEPS BETWEEN CONSECUTIVE TIME STEPS

The canonical variates $\mathbf{Z}_t$ (Fig. s9) resemble the Isomap components, as indicated by the SCSq (Fig. s10). The canonical variates $\mathbf{Z}_{t+1}$ exhibit the same patterns as $\mathbf{Z}_t$ only delayed by one time step, and are not shown.

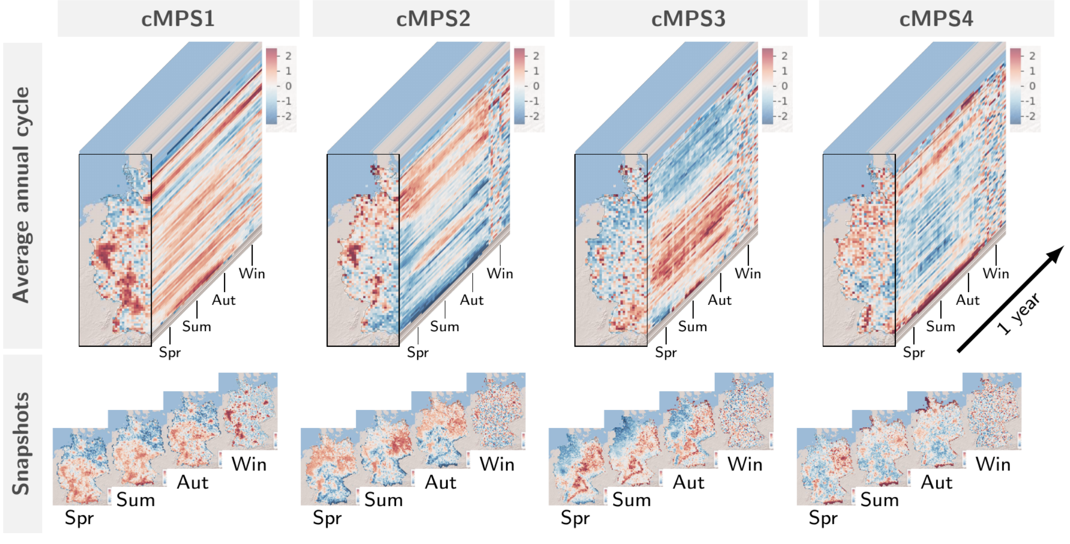

FIGURE s9 FI canonical variates $\mathbf{Z}_t$ obtained from CCorA($\mathbf{Y}_t, \mathbf{Y}_{t+1}$) with components $i = 1 \ldots 5$. The canonical variates are a linear combination of FI MEPs $\mathbf{Y}_t$. Similar colours indicate similar group behaviour.

S7 | STRUCTURAL CORRELATIONS SQUARED

The SCSq quantifies the proportion each MEP contributes to the respective canonical variates between consecutive time steps. Per time step we analysed the MEP contributions to the common patterns between consecutive time steps, CCorA$(\mathbf{Y}_t, \mathbf{Y}_{t+1})$ (Fig. s10). The SCSq indicates that each common pattern is dominated by the corresponding Isomap component. In Fig. s10(a) the SCSq of the first CV depends solely on the first MEP, while also not contributing to other CVs. The SCSq of other CVs (Fig. s10) are dominated by the corresponding MEPs during the growing season. During the non-growing season the contribution is mixed and no distinct pattern can be gauged. This is probably due to the lack of coherent patterns.

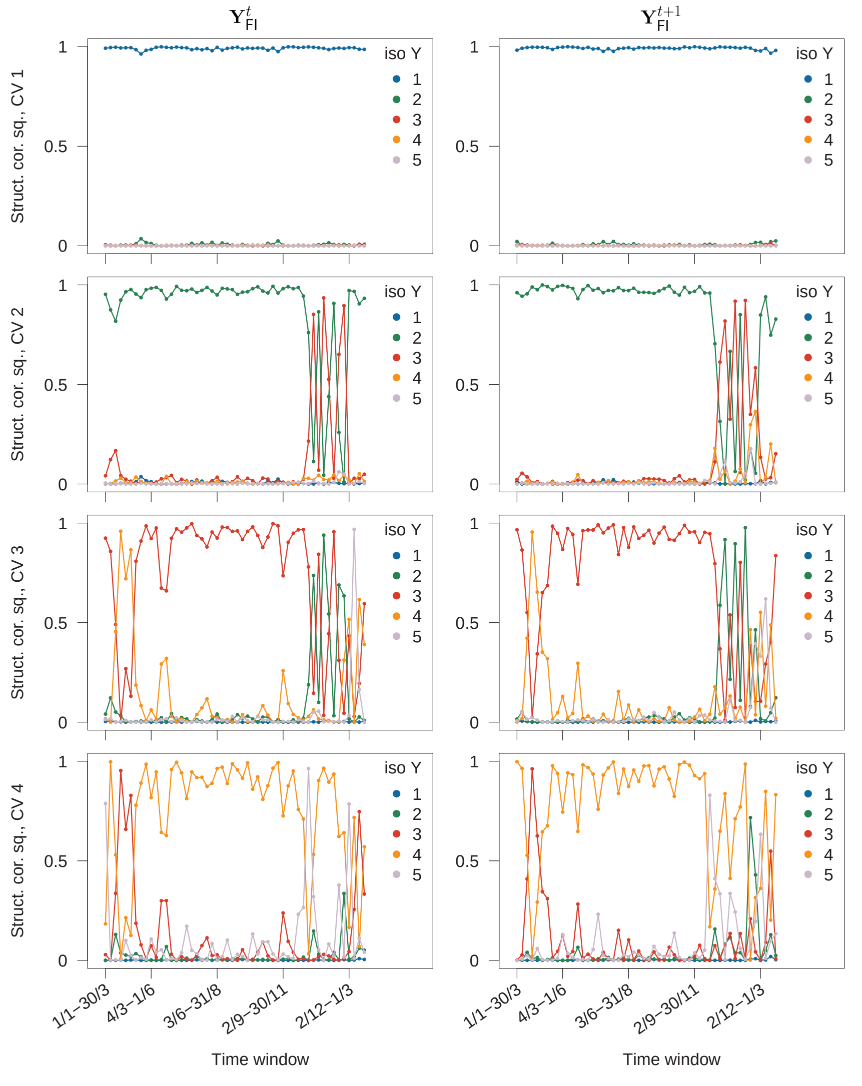

FIGURE s10 Structural correlations squared (Struct. cor. sq.) indicate the proportional contribution of each MPS $\mathbf{Y}_t, \mathbf{Y}_{t+1}$ to the respective CV per time step t. Hence, CVs are predominantly associated with the corresponding MPS $\mathbf{Y}_t, \mathbf{Y}_{t+1}$ per time step t.

S8 | CANONICAL CORRELATIONS VS. OBSERVATION COUNTS

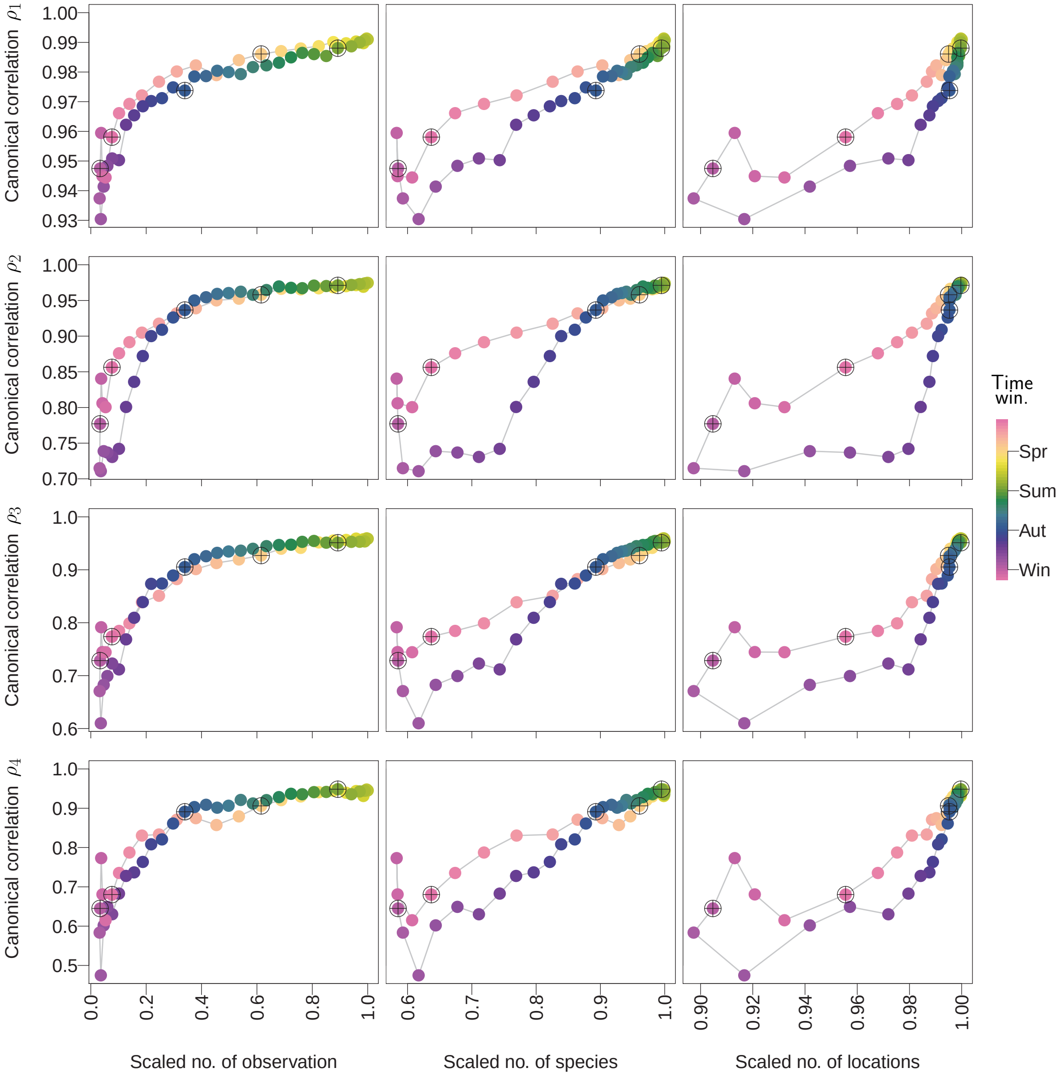

FIGURE s11 As the scaled number of observations increases, the changes in MPS between consecutive time steps decrease, as indicated by increasing and plateauing canonical correlations, $\rho_1, \rho_2, \rho_3, \rho_4$. The transition from winter to summer and vice versa are very similar. As the scaled number of locations decreases few changes in MPS can be observed from spring to autumn, as indicated across all canonical correlations $\rho_1, \rho_2, \rho_3, \rho_4$.