S1 Separation between the calibration and the evaluation data

The 91 sites of the dataset span a broad range of temperature and accumulation rate conditions (Table S1 and Fig. S1). As explained in the main text, our objective is to select the evaluation data (22 cores) randomly but still making it representative of (i) all climatic conditions and (ii) the ratio of GrIS to AIS sites of the dataset. We separate the 91 observed cores in three tiers of lowest, middle and highest T_{av} and we select randomly 7 cores in each tier for the evaluation data. We repeat this random selection until 5 to 10 out of the 21 cores are from GrIS, with the remainder from AIS. Finally, our dataset includes two sites that are climatic outliers with respect to the others (DML and spencer4 in Table S1) with high T_{av} and $\dot{b}$ values (Figure S1). We select randomly one of these for the evaluation data. Proceeding to the selection based on $\dot{b}$ rather than T_{av} would be similar given the strong correlation between both variables.

S2 The likelihood function, Eq. (8)

The covariance matrices Σ_{15} and Σ_{pc} that appear in Eq. (8) are diagonal matrices with the site-specific variances on the diagonal. At each site, we take 5% of the observed $DIP15$ and 10% of the observed $DIPpc$ as the standard deviation, and the variance value is the square of the standard deviation. We take the higher value of 10% for $DIPpc$ because density errors propagate in firn models. Equation (3) shows that densification rates depend on the density value itself, resulting in error propagation through time. As such, if a model shows an offset compared to observations at 15 m depth, it is likely to show an even stronger offset at z_{pc}. Taking a higher variance alleviates the strength of this effect on the likelihood calculations by allowing a larger spread of model results compared to observed $DIPpc$ values. The form of Eq. (8) corresponds to a normal likelihood function. This assumes that model DIP results are normally distributed around the observed values. To support this assumption, we conducted a preliminary verification of errors in $DIP15$ $(X_{15} - Y_{15})$ and $DIPpc$ $\left(X_{pc} - Y_{pc}\right)$ computed with the three original models (HL, Ar, LZ) on the entire dataset and we compute a basic Kolmogorov-Smirnov test for all six sets of errors. The resulting p-values are all above 0.05. The distribution of these errors are thus in line with a normal distribution, and this is despite the presence of some outliers in the errors in $DIP15$ that heavily disfavour the assumption of normality. We show the Quantiles-Quantiles plots for all six sets of errors in Figure S2. As explained in the main text, the form of Eq. (8) also assumes independence between errors in $DIP15$ and $DIPpc$, which is the reason why $DIPpc$ is calculated only from depths below 15 m. As such, observations-model discrepancies are essentially governed by parameter values of stage-1 densification for $DIP15$ and by parameter values of stage-2 densification for $DIPpc$, with little interaction between both. The same preliminary verification as mentioned above allows us to evaluate the correlation between $DIP15$ and $DIPpc$ errors for all three original models on the entire dataset. This yields a correlation coefficient of 0.40. The three correlation coefficients computed for each of the original models individually are all below 0.3.

S3 Convergence diagnostics

For convergence of the RWM algorithm, the chain must traverse between the peaks of the target posterior distribution multiple times. Simply examining the trace of the RWM algorithm for each parameter provides an effective way to verify this criterion. The trace is the history of accepted parameter values over the entire chain. We show this sampling history in Fig. S3. The fuzzy appearance for each parameter of each model indicates an efficient exploration of the parameter space as the samples from RWM algorithm oscillate around the posterior mode.
In addition to this, we compute the Gelman-Rubin statistic, which provides a numerical test for convergence (Gelman et al., 2013). The motivation behind this test is that if each chain (run independently) converges to the same posterior distribution, then the variances within each chain should be approximately the same. For each model (HL, Ar, LZ), we launch three different chains from different initial parameter values. For each parameter of each model, we calculate the mean within sample variance W:

$$W = \frac{s_1^2 + s_2^2 + s_3^2}{3} \quad \text{(S1)}$$

where s^2 denotes the variance of an individual chain. We then calculate the between sample variance:

$$B = \frac{n}{3-1} \sum_{i=1}^{3} \left(\bar{\theta}_i - \bar{\bar{\theta}}\right)^2 \quad \text{(S2)}$$

where n denotes the number of iterations within each chain, $\bar{\theta}_i$ the mean parameter value within each chain and $\bar{\bar{\theta}}$ is the mean of $(\bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3)$. From there, the estimate of the variance of the posterior distribution is given by:

$$\hat{\sigma}^2 = \frac{n-1}{n} W + \frac{1}{n} B \quad \text{(S3)}$$

And the Gelman-Rubin statistic is defined as:

$$R = \sqrt{\frac{\hat{\sigma}^2}{W}} \quad \text{(S4)}$$

Large values of R indicate that estimates of θ values between the different chains are significantly different. With more iterations, the chains progressively converge to the same stationary distributions and the estimates of θ become similar, resulting in values of R close to 1. We reach $R < 1.1$ for all parameters, which proves adequate convergence (Gelman et al., 2013). Two parameters of the LZ model needed a larger number of iterations to reach $R < 1.1$.

S4 Compaction anomaly calculation

For each model (HL, Ar, LZ, MAP_{HL}, MAP_{Ar}, MAP_{LZ}), the calculation of compaction anomalies is the same. First, we compute at each site i the mean annual compaction rate (specific to each model) during the reference period 1960-1980, $cmp_{ref,i}^{yr}$. During this period, the firn column, and thus also compaction rates, are in steady-state. Secondly, we compute the total compaction over the transient 2000-2017 period, $cmp_{tot,i}^{00-17}$. Finally, we calculate the total compaction anomaly over 2000-2017:

$$cmp_{an,i}^{00-17} = cmp_{tot,i}^{00-17} - 17 cmp_{ref,i}^{yr} \quad \text{(S5)}$$

We compute the root mean squared difference of the compaction anomalies $cmp_{an,i}^{00-17}$ at all sites between any pair of model j and k, $rmsd(j,k)$. We take this as an approximate average discrepancy between models j and k in compaction anomalies over GrIS, to which corresponds a discrepancy in estimated mass of ice. We take an approximate GrIS accumulation area of 1.5 10^{12} m^2 (Noël et al., 2019). Using an ice density ρ_i of 917 kg m^{-3}, we reach the total ice-mass equivalent discrepancy $\|m_{an}^{tot}\|(j,k)$ corresponding to their disagreement in cmp_{an}^{00-17} values.

$$\|m_{an}^{tot}\|(j,k) = 1.5\ 10^{12}\ rmsd(j,k)\ \rho_i \quad \text{(S6)}$$

The figures of disagreement between models given in Section 4 are the inter-model differences of $\|m_{an}^{tot}\|$. The model-specific values of cmp_{an}^{00-17} at the 27 GrIS sites are given in Table S2.

S5 Posterior correlation between parameters

The joint posterior distributions for the parameters of each model also allow us to analyse the models' internal structure, i.e. the correlation between their different parameters. The full correlation matrices are given in Fig. S4. In HL, the strongest correlation coefficients r are unsurprisingly found for the pairs of pre-exponential factor and activation energy governing densification in stage-1 (k_0^* and E_0) and in stage-2 (k_1^* and E_1) with r of 0.94 in both cases. Higher activation energies (E_0 and E_1) imply stronger thermal barriers to the densification process and thus slower densification, and the pre-exponential factors (k_0^* and k_1^*) counter-balance the effect to still match observed DIP values. In the same way, the activation energies are negatively correlated with their respective accumulation rate exponent (a and b), but more moderately (r values of approximately -0.4). The negative correlation of -0.44 between a and b themselves might be linked to the density at 15 m being the lower boundary and the upper boundary condition for the calculation of $DIP15$ and $DIPpc$ respectively. Higher values of a tend to cause lighter firn at 15 m depth. Lower E_0 values compensate for this effect on $DIP15$ because the shallow firn densifies faster due to its greater sensitivity to temperature variations. The lighter 15 m depth density also affects $DIPpc$, and lower values of b compensate for this by enhancing the densification rate, which explains the negative correlation between a and b. In Ar, the interpretation is more challenging due to the use of a same activation energy in both stages. There is a strong correlation between the activation energy E_g and both pre-exponential factors k_0^{Ar} (r = -0.94) and k_1^{Ar} (r = -0.95), for the same reason as in HL. As such, this induces a strong positive correlation between the latter parameters (r = 0.88). The negative correlation between α and k_1^{Ar} (r = -0.37) is more surprising because these parameters apply to different stages, but it reveals an interesting pattern. Higher temperatures raise densification rates at warmer sites, where accumulation rates are also higher thus further amplifying this effect. Higher accumulation rates nevertheless cause light recently deposited firn to be buried rapidly, which may cause lower density firn governed by the fast stage-1 densification to extend below 15 m. To avoid underestimation of $DIPpc$ at such sites, stage-1 densification rates must remain low enough but there is no possibility for adjusting a stage-1 specific activation energy. Lower α values generate this effect while only marginally affecting densification at colder low-accumulation sites. Thus, high k_1^{Ar} without a complementary lower α would cause $DIPpc$ underestimation at warm and high accumulation sites. Analysis of correlation coefficients in LZ is less straightforward because its governing equations, Eq. (6), are less interpretable than the plain Arrhenius relationship of HL and Ar. Still, we highlight some correlated pairs of parameters. As could be expected from Eq. (6), lz_a and lz_b are negatively correlated (r = -0.73). Also, the independent term of stage-1 densification lz_{11} is strongly correlated with the corresponding temperature-related parameter (lz_{13}, r = 0.9). The same is valid for stage-2 densification between lz_{21} and lz_{23} (r = 0.93). The positive correlation between lz_{12} and lz_{22} (r = 0.6) is discussed in the main text.

Tables

Site	Lat	Lon	Core depth [m]	Year	Mean $\dot{b}$ [m w.e.. yr^{-1}]	Mean T [°C]	ρ_0 [kg/m^3]	$DIP15$ [m]	Var $DIP15$ [m^2]	$DIPpc$ [m]	Var $DIPpc$ [m^2]
EGRIP	75.63	-35.98	20.1	2017	0.113	-29.0	285	7.816	0.153	/	/
Summit *	72.58	-38.47	22.1	2017	0.205	-28.4	330	7.500	0.141	/	/
id359	73.94	-37.63	102.4	1993	0.124	-28.8	240	6.708	0.112	11.456	1.312
id369	75.00	-30.00	19.9	1997	0.135	-27.6	335	7.454	0.139	/	/
id373	75.25	-37.62	100.8	1993	0.106	-29.5	275	7.826	0.153	12.372	1.531
id385	76.00	-43.49	109.8	1995	0.124	-29.3	315	7.857	0.154	13.186	1.739
id423 *	76.62	-36.40	143.2	1993	0.093	-29.1	310	7.716	0.149	10.666	1.138
id514	77.25	-49.22	119.6	1995	0.163	-28.3	300	7.575	0.143	13.217	1.747
id531 *	77.45	-51.06	75.0	2009	0.198	-27.4	320	7.434	0.138	/	/
id534	80.00	-41.14	96.0	1994	0.105	-28.4	335	7.811	0.153	11.345	1.287
Basin8	69.80	-36.49	29.8	2003	0.350	-25.6	300	7.396	0.137	/	/
D2	71.80	-46.34	101.3	2003	0.421	-23.4	370	7.051	0.124	14.097	1.987
D4	71.39	-43.94	143.9	2003	0.390	-24.6	300	7.394	0.137	12.770	1.631
HumboldtM *	78.47	-56.98	141.9	1995	0.384	-24.8	280	8.062	0.162	10.947	1.198
NASAE1 *	74.98	-29.97	19.9	1997	0.135	-27.6	340	7.394	0.137	/	/
spencer6 *	72.57	-37.62	82.3	1994	0.176	-29.0	360	4.889	0.060	/	/
spencer16	71.75	-40.75	15.0	1954	0.289	-27.0	340	7.216	0.130	/	/
spencer17	77.95	-39.18	60.0	1973	0.080	-29.3	300	5.002	0.063	7.781	0.605
spencer66 *	70.75	-35.96	109.0	1987	0.247	-27.3	300	7.340	0.135	14.852	2.206
spencer67	70.63	-35.83	128.6	1988	0.262	-27.0	325	7.098	0.126	14.114	1.992
spencer68 *	70.65	-35.48	105.6	1988	0.252	-27.1	325	7.172	0.129	14.505	2.104
spencer69	70.67	-35.80	24.8	1988	0.248	-27.2	305	7.184	0.129	/	/
spencer70	70.63	-35.63	100.1	1988	0.264	-26.9	290	6.772	0.115	14.026	1.967
spencer71	71.76	-35.87	77.8	1988	0.203	-28.2	275	7.043	0.124	13.094	1.715
spencer72	71.48	-35.88	25.7	1988	0.207	-28.0	330	7.223	0.130	/	/
spencer73	71.15	-35.85	70.8	1988	0.214	-27.7	340	7.230	0.131	/	/
spencer74	70.85	-35.85	26.2	1988	0.264	-26.9	330	7.087	0.126	/	/
SouthPole	-90.00	0.00	122.9	2001	0.055	-47.8	325	7.613	0.145	22.312	4.978
Newall	-77.58	162.50	111.1	1989	0.043	-31.2	305	7.160	0.128	4.132	0.171
Berkner *	-79.61	-45.72	178.2	1995	0.124	-28.3	345	6.255	0.098	9.658	0.933
DML *	-71.41	-9.92	78.2	2007	0.902	-20.6	410	6.037	0.091	10.228	1.046
id9	-76.77	-101.74	111.6	2012	0.313	-24.7	470	6.194	0.096	12.119	1.469
id10	-76.95	-121.22	62.0	2011	0.213	-28.4	355	6.947	0.121	/	/
id11	-77.06	-89.14	114.5	2001	0.346	-26.5	415	5.879	0.086	11.201	1.255

id12	-77.61	-92.25	67.8	2001	0.301	-27.8	350	6.019	0.091	/	/
id13	-77.68	-124.00	59.3	2000	0.155	-28.2	350	6.411	0.103	/	/
id14	-77.76	153.38	97.1	2006	0.048	-44.6	360	6.833	0.117	17.516	3.068
id15 *	-77.84	-102.91	70.7	2001	0.486	-25.1	415	5.853	0.086	/	/
id17	-77.88	158.46	98.5	2006	0.058	-41.1	350	6.419	0.103	11.687	1.366
id18	-77.96	-95.96	57.4	2010	0.354	-28.0	335	6.752	0.114	/	/
id19	-78.08	-120.08	57.8	2000	0.171	-27.7	315	6.253	0.098	/	/
id20	-78.12	-95.65	70.5	2001	0.324	-27.7	385	6.265	0.098	/	/
id22 *	-78.33	-124.48	59.9	2000	0.152	-27.7	285	6.509	0.106	8.989	0.808
id24	-78.43	-115.92	59.8	2000	0.318	-27.8	390	6.295	0.099	/	/
id26	-78.73	-111.50	60.7	2000	0.329	-27.8	350	6.427	0.103	/	/
id28	-79.04	149.68	100.1	2006	0.040	-44.6	405	6.703	0.112	15.584	2.428
id29 *	-79.13	-122.27	63.1	2000	0.127	-27.8	300	6.507	0.106	9.926	0.985
id30	-79.16	-104.97	72.7	2001	0.306	-28.7	400	5.921	0.088	/	/
id33	-79.38	-111.24	104.8	2000	0.239	-28.2	370	6.159	0.095	12.943	1.675
id35 *	-79.48	-112.09	160.0	2011	0.162	-28.0	460	6.181	0.096	11.824	1.398
id39	-80.62	-122.63	57.5	1999	0.094	-25.9	370	6.253	0.098	/	/
id43	-81.20	-126.17	48.3	1999	0.070	-24.5	325	6.268	0.098	4.975	0.247
id46	-82.00	-110.01	62.2	2002	0.180	-27.8	340	6.161	0.095	/	/
id48	-83.50	-104.99	61.7	2002	0.220	-31.0	360	6.098	0.093	/	/
id49 *	-84.40	140.63	50.1	2007	0.023	-45.4	340	6.886	0.119	/	/
id50	-85.00	-105.00	44.9	2002	0.157	-36.3	360	6.422	0.103	/	/
id51	-85.78	145.72	41.7	2007	0.033	-46.1	310	6.767	0.114	/	/
id52 *	-86.50	-107.99	71.6	2002	0.147	-38.8	340	6.882	0.118	/	/
id53	-86.84	95.31	20.8	2003	0.042	-53.3	355	6.535	0.107	/	/
id54 *	-88.00	-107.98	54.1	2002	0.133	-41.4	355	7.009	0.123	/	/
id55	-88.51	178.53	99.3	2007	0.081	-48.2	320	6.880	0.118	/	/
id56	-89.93	144.39	139.5	2002	0.080	-48.6	345	6.319	0.100	25.046	6.273
spencer1	-80.00	-120.00	307.0	1968	0.120	-27.2	350	6.987	0.122	10.314	1.064
spencer4	-66.72	113.18	200.9	1989	1.060	-22.0	380	7.848	0.154	12.847	1.650
spencer5	-74.50	123.17	49.5	1980	0.037	-51.8	345	8.262	0.171	/	/
spencer7	-85.25	166.50	79.9	19997	0.028	-39.7	305	7.003	0.123	8.202	0.673
spencer8	-66.77	112.80	180.0	1997	0.488	-22.7	385	7.385	0.136	10.640	1.132
spencer22	-73.60	-12.43	25.5	1996	0.220	-22.5	380	3.920	0.038	/	/
spencer25	-74.02	-12.02	26.5	1996	0.171	-30.7	390	5.412	0.073	/	/
spencer29 *	-75.00	2.00	20.6	1996	0.072	-42.9	320	7.602	0.144	/	/
spencer33	-70.68	44.32	123.5	1978	0.114	-33.1	385	6.385	0.102	7.022	0.493

spencer34 *	-70.68	44.32	109.0	1978	0.114	-33.1	375	6.161	0.095	6.909	0.477
spencer61	-73.10	39.75	99.7	1978	0.069	-42.3	360	7.005	0.123	16.245	2.639
spencer62 *	-71.18	45.97	100.2	1997	0.091	-38.2	395	7.049	0.124	16.344	2.671
spencer76	-90.00	0.00	122.1	1997	0.055	-47.8	360	4.906	0.129	/	/
spencer77	-75.00	147.00	15.8	1961	0.042	-46.1	385	7.184	0.130	/	/
spencer78 *	-74.00	143.00	16.0	1961	0.043	-45.5	375	7.205	0.128	/	/
spencer79	-73.00	142.00	15.7	1961	0.057	-44.0	325	7.148	0.118	/	/
spencer80	-73.00	141.00	16.0	1961	0.057	-44.0	355	6.876	0.120	/	/
spencer81	-72.00	140.00	16.9	1961	0.080	-42.7	335	6.936	0.117	/	/
spencer82 *	-71.00	139.00	15.6	1961	0.120	-41.6	375	6.848	0.115	/	/
spencer83	-72.00	143.00	15.7	1961	0.087	-41.3	405	6.796	0.118	/	/
spencer84	-72.00	146.00	16.2	1961	0.086	-40.9	410	6.876	0.114	/	/
spencer85	-72.00	148.00	15.9	1961	0.096	-40.2	360	6.745	0.121	/	/
spencer86	-72.00	151.00	15.8	1961	0.103	-39.7	400	6.963	0.103	/	/
spencer87	-72.00	154.00	15.9	1961	0.130	-38.0	355	6.430	0.124	/	/
spencer88	-72.00	156.00	15.7	1961	0.130	-37.6	395	7.050	0.111	/	/
spencer89	-72.00	159.00	15.7	1961	0.115	-35.7	370	6.665	0.139	10.046	1.009
spencer90	-83.47	138.80	340.5	1994	0.020	-45.2	420	7.461	0.124	3.530	0.125
spencer91	-83.47	-138.80	47.0	1987	0.058	-27.1	295	7.037	0.193	20.368	4.148
spencer92	-78.47	106.80	179.3	1996	0.022	-54.6	360	8.790	0.153	/	/

Table S1. The 91 firn core dataset used in this study. * symbols indicate the core is part of the evaluation data. Lat and Lon designate latitude and longitude respectively. Year indicates the year of drilling of the core. $\dot{b}$ is the accumulation rate. T is the temperature. ρ_0 is the surface density boundary condition that was derived individually for each core by extrapolating density measurements until the surface. Var designates the site-specific variance used for the terms of Σ_{15} and Σ_{pc} (see Text S2 for their calculation).

cmp_{an}^{00-17} [m]	HL	MAP_{HL}	Ar	MAP_{Ar}	LZ	MAP_{LZ}
EGRIP	0.626	0.648	0.735	0.656	0.675	0.695
Summit	0.464	0.461	0.442	0.442	0.481	0.475
id359	0.748	0.763	0.866	0.781	0.815	0.837
id369	0.609	0.587	0.597	0.574	0.623	0.617
id373	0.613	0.637	0.735	0.649	0.664	0.693
id385	0.319	0.333	0.407	0.356	0.373	0.393
id423	0.528	0.559	0.649	0.566	0.565	0.600
id514	0.358	0.367	0.409	0.378	0.409	0.423
id531	0.336	0.333	0.304	0.328	0.373	0.371
id534	0.216	0.214	0.189	0.205	0.229	0.237
Basin8	0.764	0.760	0.728	0.671	0.641	0.611
D2	0.244	0.255	0.222	0.164	0.158	0.133
D4	0.560	0.553	0.550	0.465	0.500	0.479
HumboldtM	0.630	0.616	0.606	0.534	0.562	0.545
NASAE1	0.591	0.568	0.578	0.555	0.605	0.596
spencer6	0.380	0.386	0.391	0.367	0.393	0.399
spencer16	0.499	0.489	0.500	0.428	0.462	0.448
spencer17	0.503	0.534	0.611	0.538	0.536	0.566
spencer66	0.963	0.944	0.955	0.912	0.926	0.889
spencer67	0.832	0.812	0.821	0.762	0.761	0.744
spencer68	0.895	0.875	0.895	0.831	0.838	0.811
spencer69	0.935	0.915	0.920	0.880	0.895	0.857
spencer70	1.078	1.056	1.060	1.017	1.025	0.982
spencer71	1.094	1.066	1.126	1.042	1.107	1.093
spencer72	0.809	0.806	0.852	0.780	0.812	0.799
spencer73	0.793	0.786	0.835	0.756	0.782	0.763
spencer74	0.856	0.842	0.861	0.785	0.786	0.771

Table S2. Total compaction anomaly over 2000-2017 (cmp_{an}^{00-17}) at the 27 GrIS sites. See text S3 for calculation details.

Figures

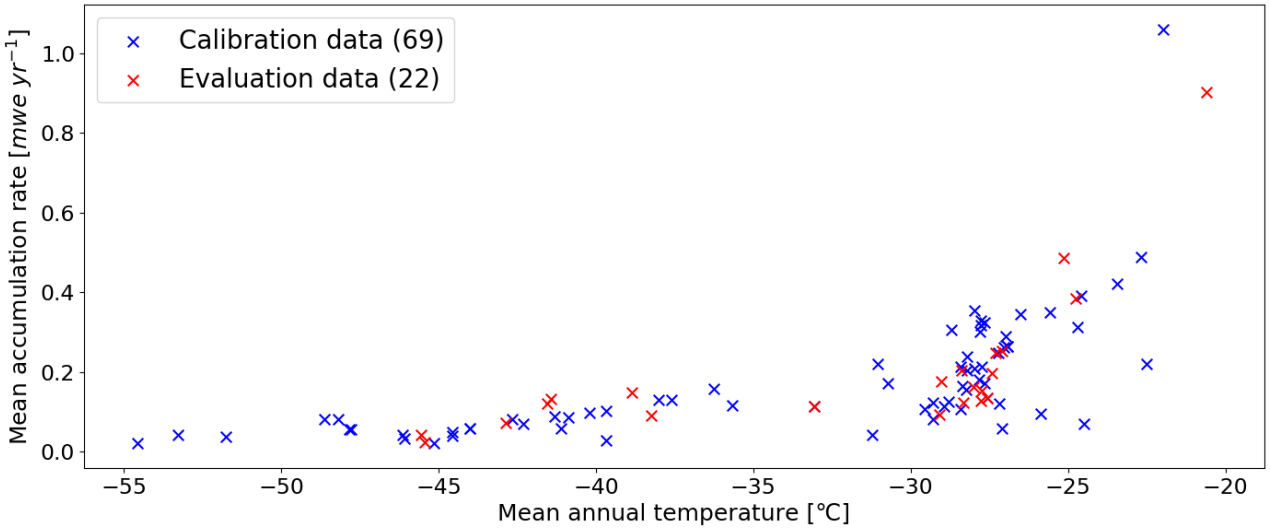

Figure S1. Climatic conditions at the 91 sites of the dataset

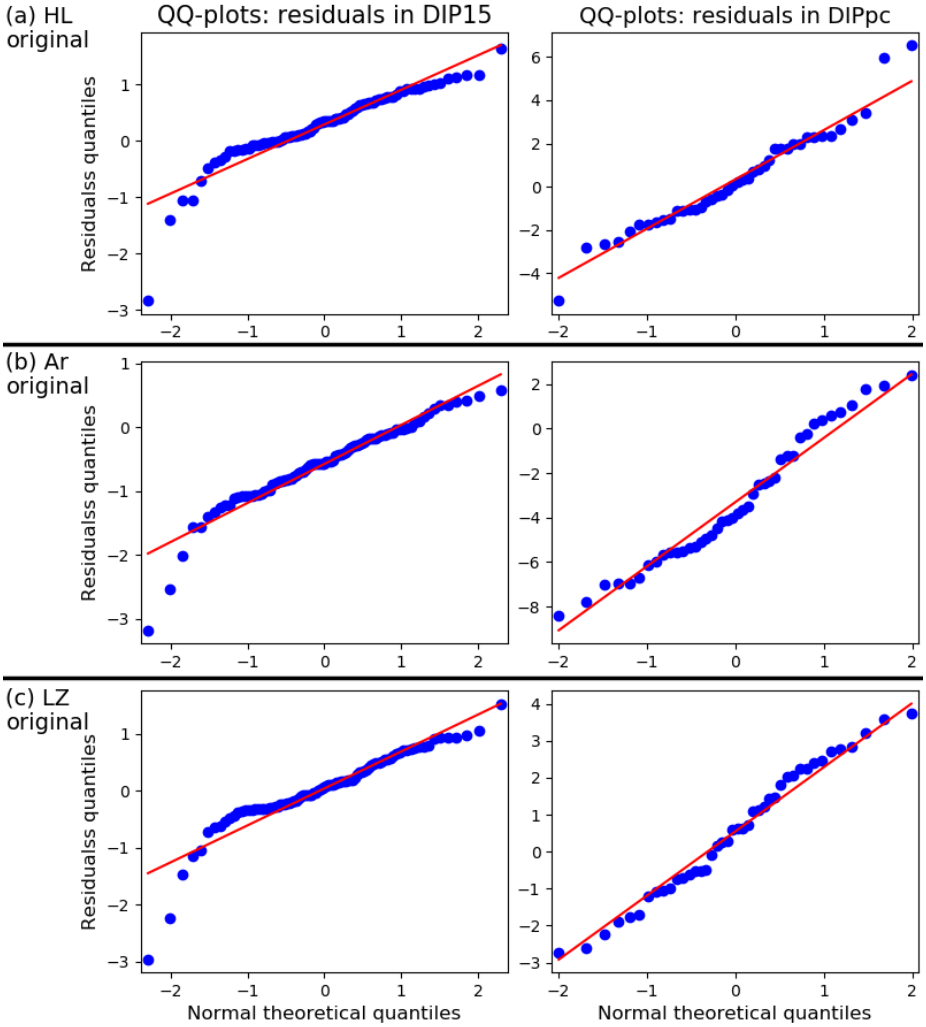

Figure S2. Quantiles-Quantiles plots for the errors of the original models computed on the entire dataset. The alignment of the points along the red line informs about the fit to a normal distribution.

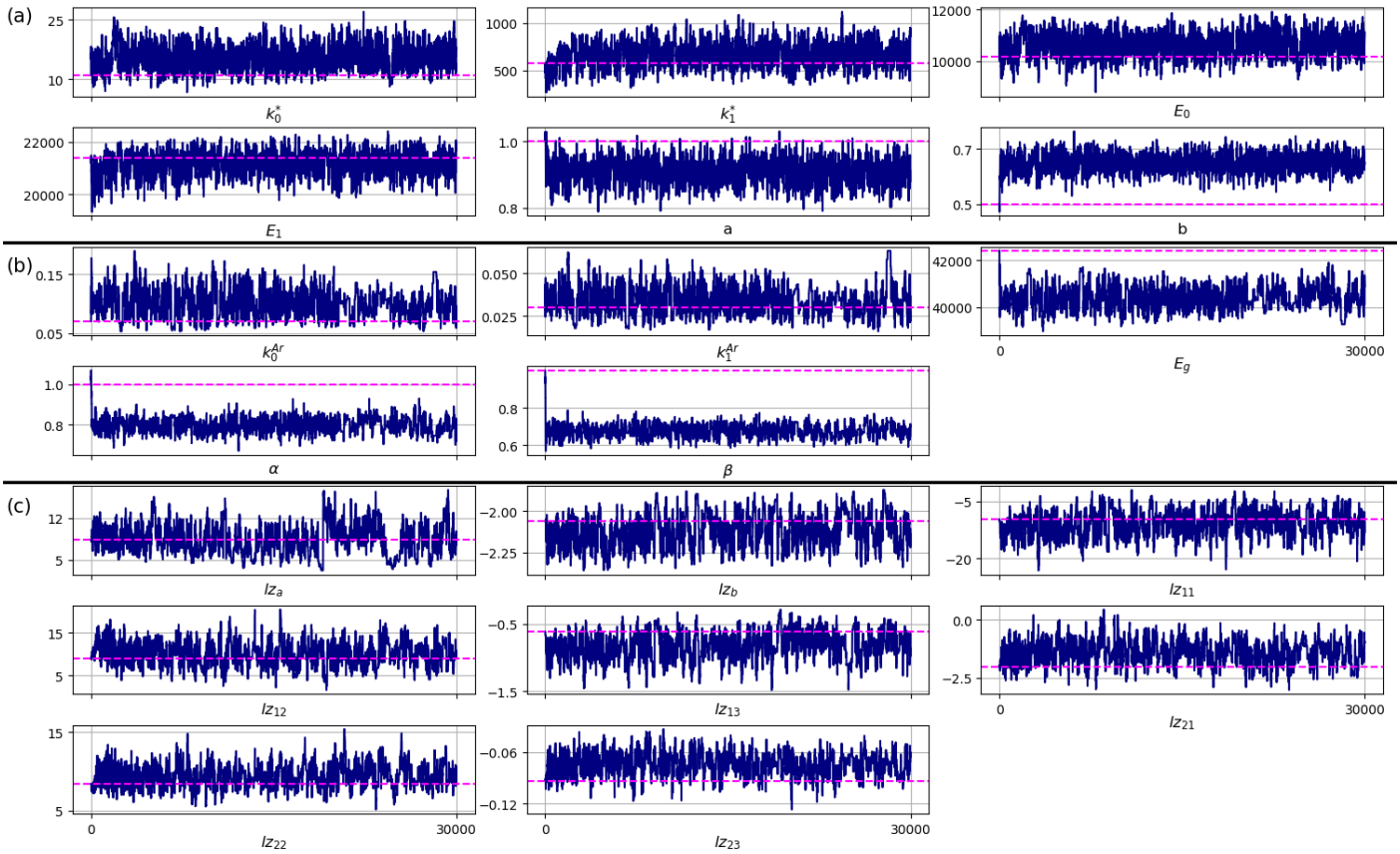

Figure S3. Sampling chains of each parameter for (a) HL, (b) Ar, (c) LZ. The x-axis displays the iteration number, the y-axis displays the parameter value. The dashed pink line shows the value of the original model, which is also the starting point of each chain.

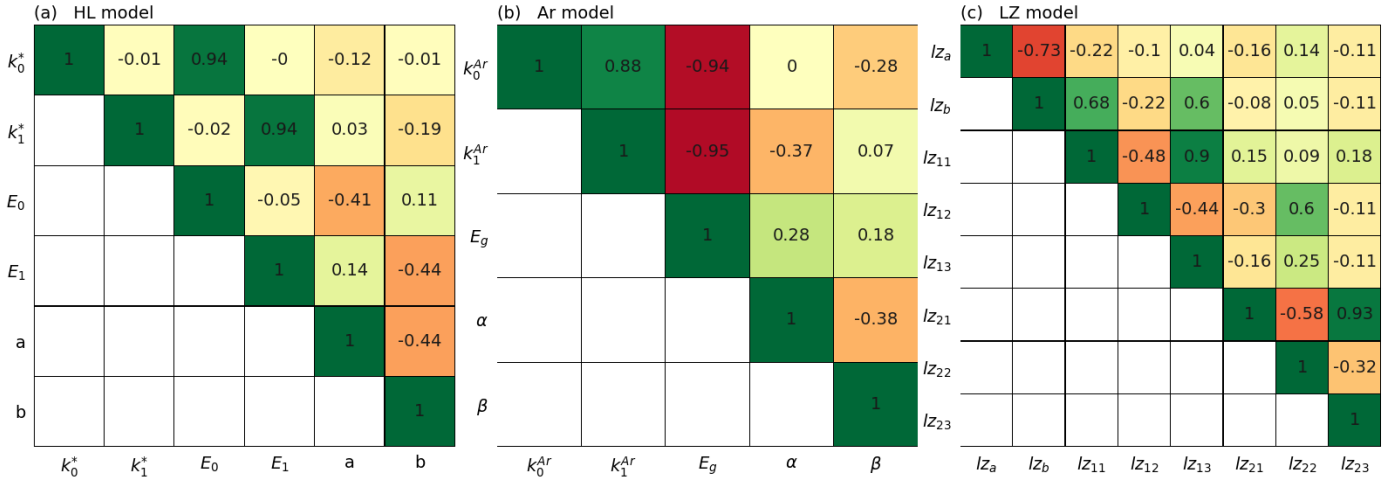

Figure S4. Posterior correlation matrices.