

Supplement of

Rediscovering the climate analogue method: insights from a data-driven approach

P.-C. Chen et al.

S1. Summary of model combinations

All model combinations compared for the quality of the resulting analogues are summarised in Table S1. As seen, the traditional Euclidean distance (ED), the original AtmoDist (AD) and the proposed ClimaDist are considered. In terms of input data, we tested on using only temperature (t) and all selected variables –temperatures (t), u- and v-wind components (u, v) and geopotential heights (z), and on two different input data structures. These include in the original time point, aggregation and concatenation (see Sect. 3.2.2 for details). In terms of distance metrics, in addition to the original Euclidean distance (d_{ED}^O and d_{ED} , applied to the original value space and to the feature space, respectively), the image Euclidean distance (IMED, denoted d_{IMED}) and the structural similarity index metric (SSIM, denoted d_{SSIM}) are explored.

Table S1: All combinations of models and input data settings tested in this study. (Note: d_{ED}^O is the distance metric proposed by Delle Monache et al. (2011), applied to the original data space.)

	Method	Input data setting	Distance measures	Variables
AD_t	AD	time points	d_{ED}	t
AD_{tuvz}	AD	time points	d_{ED}	t, u, v, z
ED_t	ED	time points	d_{ED}^O	t
ED_{tuvz}	ED	time points	d_{ED}^O	t, u, v, z
$CD_t^{a,imed}$	CD	Aggregation	d_{IMED}	t
$CD_t^{a,ssim}$	CD	Aggregation	d_{SSIM}	t
$CD_{tuvz}^{a,imed}$	CD	Aggregation	d_{IMED}	t, u, v, z
$CD_{tuvz}^{a,ssim}$	CD	Aggregation	d_{SSIM}	t, u, v, z
$CD_t^{c,imed}$	CD	Concatenation	d_{IMED}	t
$CD_t^{c,ssim}$	CD	Concatenation	d_{SSIM}	t
$CD_{tuvz}^{c,imed}$	CD	Concatenation	d_{IMED}	t, u, v, z
$CD_{tuvz}^{c,ssim}$	CD	Concatenation	d_{SSIM}	t, u, v, z

S2. Results of analogue determination from all model combinations

The results obtained from all model combinations (see Table S1) are summarised here. For the quality of the best individual analogues, the results of 30-day means of the selected atmospheric variables are provided with the Perkins' skill scores (S_{score}), Spearman correlation coefficients (ρ) and regression coefficients (β) (see Table S2). In addition, the results of capturing 30-day variability, maximum and minimum of temperatures from best individual analogues are given in Table S3.

For best ensemble analogues, the results of 30-day means of the selected atmospheric variables are provided in Table S4, in terms of probabilistic Nash–Sutcliffe Efficiency coefficients (NS^{prob}) and the corresponding average and variance terms. In addition, the results of capturing 30-day variability, maximum and minimum of temperatures from best ensemble analogues are given in Tables S5 and S6.

Table S2: Performance metrics of the 30-day means of the selected atmospheric variables of the best individual analogues determined from all model combinations. Selected variables include: temperature (T_{mean}), u-wind (U_{mean}), v-wind (V_{mean}), and geopotential (Z_{mean}), and the metrics include Perkins’ skill score (S_{score}), Spearman correlation coefficient (ρ), and regression coefficient (β).

	Perkins’ skill score (S_{score})				Correlation coefficient (ρ)				Regression coefficient (β)			
	T_{mean}	U_{mean}	V_{mean}	Z_{mean}	T_{mean}	U_{mean}	V_{mean}	Z_{mean}	T_{mean}	U_{mean}	V_{mean}	Z_{mean}
AD_t	0.917	0.903	0.833	0.861	0.082	0.320	-0.208	0.156	0.090	0.289	-0.245	0.057
AD_{tuvz}	0.917	0.833	0.889	0.861	0.158	0.145	0.104	0.122	0.132	0.094	0.095	0.250
ED_t	0.861	0.806	0.972	0.736	0.885	0.153	0.109	0.564	0.934	0.191	0.208	0.782
ED_{tuvz}	0.903	0.889	0.903	0.889	0.902	0.799	0.777	0.840	0.935	0.850	0.936	0.828
$CD_t^{a,imed}$	0.986	0.931	0.861	0.889	0.999	0.424	0.327	0.685	0.992	0.502	0.323	0.694
$CD_t^{a,ssim}$	0.972	0.736	0.903	0.819	0.998	0.420	0.479	0.763	0.992	0.532	0.526	0.787
$CD_{tuvz}^{a,imed}$	0.931	0.958	0.917	0.972	0.964	0.875	0.882	0.968	0.940	0.963	0.943	0.973
$CD_{tuvz}^{a,ssim}$	0.958	0.917	0.944	0.944	0.992	0.875	0.867	0.947	1.013	0.930	0.962	1.004
$CD_t^{c,imed}$	0.944	0.806	0.931	0.931	0.988	0.171	0.016	0.662	0.960	0.144	0.060	0.688
$CD_t^{c,ssim}$	0.931	0.958	0.764	0.764	0.971	0.327	0.204	0.516	0.942	0.330	0.286	0.550
$CD_{tuvz}^{c,imed}$	0.958	0.806	0.875	0.875	0.991	0.366	-0.027	0.681	0.965	0.391	-0.054	0.706
$CD_{tuvz}^{c,ssim}$	0.944	0.778	0.750	0.847	0.979	0.184	0.315	0.571	0.969	0.216	0.361	0.639

Table S3: Performance metrics of the 30-day mean (T_{mean}), standard deviation (T_{std}), minimum (T_{min}), and maximum (T_{max}) temperatures of the best individual analogues determined from all model combinations.

	Perkins' skill score (S_{score})				Correlation coefficient (ρ)				Regression coefficient (β)			
	T_{mean}	T_{std}	T_{max}	T_{min}	T_{mean}	T_{std}	T_{max}	T_{min}	T_{mean}	T_{std}	T_{max}	T_{min}
AD_t	0.917	0.847	0.889	0.875	0.091	0.284	0.044	0.103	0.090	0.212	0.045	0.093
AD_{tuvz}	0.819	0.819	0.833	0.889	-0.059	-0.169	-0.004	-0.067	-0.030	-0.126	0.025	-0.063
ED_t	0.861	0.917	0.889	0.944	0.867	0.510	0.779	0.806	0.934	0.585	0.822	0.805
ED_{tuvz}	0.903	0.917	0.944	0.958	0.902	0.629	0.842	0.833	0.935	0.574	0.889	0.839
$CD_t^{a,imed}$	0.986	0.875	0.889	0.931	0.999	0.569	0.876	0.924	0.992	0.394	0.894	0.848
$CD_t^{a,ssim}$	0.972	0.792	0.819	0.931	0.998	0.488	0.886	0.874	0.992	0.382	0.936	0.809
$CD_{tuvz}^{a,imed}$	0.931	0.875	0.931	0.833	0.964	0.569	0.888	0.864	0.944	0.394	0.959	0.722
$CD_{tuvz}^{a,ssim}$	0.958	0.875	0.958	0.875	0.981	0.578	0.894	0.887	0.984	0.484	0.944	0.877
$CD_t^{c,imed}$	0.944	0.792	0.889	0.861	0.991	0.561	0.954	0.871	0.960	0.544	0.973	0.798
$CD_t^{c,ssim}$	0.931	0.889	0.875	0.861	0.977	0.645	0.891	0.859	0.970	0.534	0.933	0.802
$CD_{tuvz}^{c,imed}$	0.958	0.833	0.958	0.875	0.991	0.619	0.944	0.859	0.965	0.574	0.944	0.765
$CD_{tuvz}^{c,ssim}$	0.944	0.875	0.792	0.792	0.979	0.676	0.907	0.802	0.969	0.653	0.971	0.702

Table S4: Probabilistic Nash-Sutcliffe Efficiency coefficient (NS^{prob}), and the corresponding average (denoted NS_O^{prob} , which is equivalent to the original Nash-Sutcliffe Efficiency coefficient) and variance terms (NS_{Var}^{prob}) of 30-day mean temperature (T_{mean}), u-wind (U_{mean}), v-wind (V_{mean}) and geopotential (Z_{mean}) of the best ensemble analogues determined from all model combinations.

	NS^{prob}				NS_O^{prob}				NS_{Var}^{prob}			
	T_{mean}	U_{mean}	V_{mean}	Z_{mean}	T_{mean}	U_{mean}	V_{mean}	Z_{mean}	T_{mean}	U_{mean}	V_{mean}	Z_{mean}
AD_t	-1.814	-3.396	-1.827	-1.503	-0.598	-0.554	-0.634	-0.537	1.216	2.842	1.193	0.966
AD_{tuvz}	-1.789	-3.603	-1.635	-1.282	-0.653	-0.972	-0.574	-0.411	1.136	2.631	1.061	0.871
ED_t	0.728	-1.310	-1.201	-0.429	0.897	0.019	0.087	0.185	0.168	1.329	1.288	0.614
ED_{tuvz}	0.741	0.322	0.276	0.467	0.926	0.859	0.906	0.848	0.185	0.537	0.630	0.381
$CD_t^{a,imed}$	0.996	-0.606	-0.507	0.166	0.998	0.339	0.395	0.580	0.002	0.945	0.902	0.414
$CD_t^{a,ssim}$	0.944	0.778	0.750	0.847	0.979	0.184	0.315	0.571	0.969	0.216	0.361	0.639
$CD_{tuvz}^{a,imed}$	0.891	0.588	0.468	0.874	0.942	0.855	0.834	0.957	0.052	0.268	0.366	0.083
$CD_{tuvz}^{a,ssim}$	0.971	0.653	0.575	0.825	0.992	0.928	0.886	0.948	0.020	0.275	0.311	0.123
$CD_t^{c,imed}$	0.966	-1.082	-1.034	-0.052	0.990	0.146	0.180	0.562	0.023	1.228	1.214	0.613
$CD_t^{c,ssim}$	0.906	-1.329	-1.483	-0.266	0.962	0.092	-0.175	0.461	0.056	1.421	1.307	0.727
$CD_{tuvz}^{c,imed}$	0.971	-1.018	-1.032	-0.083	0.991	0.159	0.195	0.550	0.020	1.177	1.227	0.633
$CD_{tuvz}^{c,ssim}$	0.925	-1.228	-1.443	-0.078	0.975	0.094	-0.169	0.531	0.050	1.321	1.274	0.609

Table S5: Probabilistic Nash-Sutcliffe Efficiency coefficient (NS^{prob}) of 30-day mean (T_{mean}), standard deviation (T_{std}), minimum (T_{min}), and maximum (T_{max}) temperatures of the best ensemble analogues determined from all model combinations.

	$NS_{T_{mean}}^{prob}$	$NS_{T_{std}}^{prob}$	$NS_{T_{max}}^{prob}$	$NS_{T_{min}}^{prob}$
AD_t	-1.814	-1.619	-2.599	-2.053
AD_{tuvz}	-1.596	-1.284	-1.439	-1.252
ED_t	0.728	-0.025	0.532	0.597
ED_{tuvz}	0.741	-0.073	0.563	0.589
$CD_t^{a,imed}$	0.996	0.063	0.762	0.768
$CD_t^{a,ssim}$	0.905	0.115	0.771	0.773
$CD_{tuvz}^{a,imed}$	0.891	0.063	0.762	0.768
$CD_{tuvz}^{a,ssim}$	0.971	0.175	0.764	0.779
$CD_t^{c,imed}$	0.966	0.164	0.825	0.723
$CD_t^{c,ssim}$	0.906	0.246	0.706	0.577
$CD_{tuvz}^{c,imed}$	0.971	0.280	0.810	0.729
$CD_{tuvz}^{c,ssim}$	0.925	0.207	0.725	0.581

Table S6: The average (NS_O^{prob}) and the variance (NS_{Var}^{prob}) terms of the Probabilistic Nash-Sutcliffe Efficiency coefficient (NS^{prob}) of 30-day mean (T_{mean}), standard deviation (T_{std}), minimum (T_{min}), and maximum (T_{max}) temperatures of the best ensemble analogues determined from all model combinations.

	$NS_{O,T_{mean}}^{prob}$	$NS_{O,T_{std}}^{prob}$	$NS_{O,T_{max}}^{prob}$	$NS_{O,T_{min}}^{prob}$	$NS_{Var,T_{mean}}^{prob}$	$NS_{Var,T_{std}}^{prob}$	$NS_{Var,T_{max}}^{prob}$	$NS_{Var,T_{min}}^{prob}$
AD_t	-0.598	-0.294	-0.963	-0.591	1.216	1.326	1.636	1.462
AD_{tuvz}	-0.685	-0.287	-0.575	-0.444	0.911	0.997	0.864	0.807
ED_t	0.897	0.500	0.830	0.819	0.168	0.525	0.298	0.222
ED_{tuvz}	0.926	0.588	0.839	0.830	0.185	0.515	0.276	0.242
$pure_{imed}$	0.992	0.310	0.869	0.840	0.034	0.318	0.135	0.122
$pure_{ssim}$	0.888	0.469	0.709	0.812	0.322	0.532	0.502	0.332
$CD_t^{a,imed}$	0.998	0.519	0.897	0.882	0.002	0.456	0.134	0.114
$CD_t^{a,ssim}$	0.998	0.549	0.903	0.880	0.003	0.434	0.144	0.099
$CD_{tuvz}^{a,lzma}$	0.840	0.392	0.780	0.769	0.275	0.520	0.371	0.317
$CD_{tuvz}^{a,imed}$	0.942	0.519	0.897	0.847	0.052	0.456	0.158	0.148
$CD_{tuvz}^{a,ssim}$	0.992	0.591	0.881	0.904	0.020	0.415	0.117	0.124
$CD_t^{c,imed}$	0.990	0.528	0.922	0.868	0.023	0.363	0.097	0.145
$CD_t^{c,ssim}$	0.962	0.665	0.856	0.767	0.056	0.420	0.150	0.190
$CD_{tuvz}^{c,imed}$	0.991	0.640	0.865	0.914	0.020	0.321	0.136	0.104
$CD_{tuvz}^{c,ssim}$	0.975	0.634	0.878	0.797	0.050	0.427	0.153	0.216