Supplementary Material: An Ensemble Empirical Mode Decomposition, Self-Organizing Map, and Linear Genetic Programming Approach for Forecasting River Streamflow

Jonathan T. Barge and Hatim O. Sharif

LGP-1–LGP-3

Table S1. Average impact and frequency of occurrence for of input parameters of models that used LGP as a standalone application.

Variable	Freq. of Occ.	Average Impact	Variable	Freq. of Occ.	Average Impact
P(t)	100%	5.817%	Q(t−1)	100%	85.175%
P(t−1)	100%	1.167%	Q(t−2)	100%	40.974%
P(t−2)	20%	0.191%	Q(t−3)	83%	4.360%
P(t−3)	27%	0.034%	Q(t−4)	43%	0.863%
P(t−4)	10%	0.000%	Q(t−5)	17%	0.000%
P(t−5)	7%	0.135%	Q(t−6)	7%	0.000%
P(t−6)	0%	0.000%	Q(t−7)	7%	0.000%
P(t−7)	7%	0.000%	Q(t−8)	10%	0.000%
P(t−8)	10%	0.000%	Q(t−9)	3%	0.000%
P(t−9)	0%	0.000%	Q(t−10)	3%	0.000%

Table S2. Variables included in LGP standalone models.

Model	Input Variables
GP Int.	P(t), ..., P(t−9), Q(t−1), …, Q(t−10)
LGP 1	Q(t−1), Q(t−2), Q(t−3)
LGP 1.2	P(t), P(t−1), P(t−2)
LGP 1.3	P(t), Q(t−1)
LGP 2	P(t), Q(t−1), Q(t−2)
LGP 1.5	P(t), Q(t−1), Q(t−2), Q(t−3)
LGP 3	P(t), P(t−1), Q(t−1), Q(t−2), Q(t−3)
LGP 1.7	P(t), P(t−1), P(t−2), Q(t−1), Q(t−2), Q(t−3)

Table S3. Statistical parameters for LGP standalone models in S2.

Model		AARE	*E*	*R*	NMBE	NRMSE	%MF	E_{per}
LGP Int.	Training	22.386	0.943	0.971	−0.008	0.387	−9.932	0.666
	Validation	27.886	0.937	0.968	−0.204	0.401	17.166	0.666
LGP-1	Training	38.261	0.911	0.955	−0.022	0.485	−3.934	0.477
	Validation	38.305	0.901	0.949	0.083	0.502	−3.530	0.477
LGP 1.2	Training	346.284	0.379	0.615	−2.581	1.283	−8.690	
	Validation	372.838	0.364	0.603	−3.937	1.274	23.371	
LGP 1.3	Training	27.033	0.892	0.944	−0.842	0.535	−13.615	0.361
	Validation	28.529	0.888	0.942	−1.173	0.535	1.865	0.407
LGP-2	Training	17.223	0.943	0.971	−0.828	0.387	−12.740	0.666
	Validation	17.118	0.937	0.968	−0.905	0.401	0.156	0.666
LGP 1.5	Training	20.154	0.939	0.969	0.011	0.403	−8.141	0.638
	Validation	19.852	0.934	0.966	0.137	0.411	7.828	0.650
LGP-3	Training	22.026	0.942	0.970	0.615	0.393	−14.663	0.656
	Validation	23.360	0.942	0.970	0.312	0.386	3.959	0.691
LGP 1.7	Training	46.788	0.943	0.971	3.556	0.390	−3.206	0.886
	Validation	50.239	0.936	0.967	2.896	0.405	14.815	0.886

LGP-5

Table S4. Input variables for models depicted separated Rising/Falling trends using LGP as a standalone application.

Model		Input Variables
LGP-5.1	Rising Trends	P(*t*), Q(*t*−1), Q(*t*−2)
LGP-5.2	Rising Trends	P(*t*), P(*t*−1), Q(*t*−1), Q(*t*−2)
LGP-5.3	Falling Trends	P(*t*), Q(*t*−1), Q(*t*−2)
LGP-5.4	Falling Trends	P(*t*), P(*t*−1), Q(*t*−1), Q(*t*−2)

Table S5. Statistical parameters for models depicted separated Rising/Falling trends using LGP as a standalone application as depicted in Table S4.

Model		AARE	*E*	*R*	NMBE	NRMSE	%MF	E_{per}
LGP-5.1	Training	32.230	0.963	0.981	1.782	0.295	−11.411	0.766
	Validation	38.744	0.948	0.974	−0.603	0.350	−0.003	0.735
LGP-5.2	Training	33.625	0.967	0.983	4.307	0.260	−9.219	0.778
	Validation	43.952	0.959	0.979	1.385	0.296	9.016	0.775
LGP-5.3	Training	13.505	0.970	0.985	−2.721	0.275	1.946	0.850
	Validation	12.926	0.973	0.986	−3.084	0.247	−3.738	0.859
LGP-5.4	Training	9.065	0.971	0.985	−1.850	0.241	−2.827	0.916
	Validation	8.511	0.968	0.984	−1.974	0.236	1.590	0.907

Additional—Precipitation Error on LGP's Ability to Capture Streamflow

Table S6. Input Variables for tested error in precipitation record using LGP as a standalone application.

Model		Input Variables
LPG-P$_{err}$	50% Error	P(t), P(t−1), Q(t−1), Q(t−2), Q(t−3)
	150% Error	P(t), P(t−1), Q(t−1), Q(t−2), Q(t−3)

Table S7. Statistical parameters for tested error in precipitation record using LGP as a standalone application as depicted in Table S6.

Model		AARE	E	R	NMBE	NRMSE	%MF	E_{per}
LPG-P$_{err}$	Training	48.794	0.943	0.971	6.510	0.387	−7.046	0.666
	Validation	33.720	0.926	0.962	−2.264	0.434	−0.324	0.610
	Training	18.353	0.943	0.971	−3.011	0.389	−16.410	0.664
	Validation	22.421	0.936	0.967	2.242	0.403	1.589	0.663

LGP-8

Table S8. Frequency and average impact of input variables from models using EEMD-LGP design to capture streamflow.

Variable	Frequency	Average Impact	Variable	Frequency	Average Impact
Q(t−1)	100%	91.953%	Imf 7(t−1)	0%	0.000%
Q(t−2)	100%	83.815%	Imf 7(t−2)	3%	0.126%
Q(t−3)	17%	1.705%	Imf 7(t−3)	7%	0.000%
P(t)	73%	0.036%	Imf 8(t−1)	20%	0.279%
P(t−1)	7%	0.000%	Imf 8(t−2)	0%	0.000%
Imf 1(t−1)	100%	15.445%	Imf 8(t−3)	0%	0.000%
Imf 1(t−2)	100%	0.712%	Imf 9(t−1)	0%	0.000%
Imf 1(t−3)	33%	0.170%	Imf 9(t−2)	3%	0.022%
Imf 2(t−1)	23%	0.250%	Imf 9(t−3)	7%	0.012%
Imf 2(t−2)	100%	0.609%	Imf 10(t−1)	0%	0.000%
Imf 2(t−3)	43%	0.259%	Imf 10(t−2)	0%	0.000%
Imf 3(t−1)	20%	0.087%	Imf 10(t−3)	3%	0.000%
Imf 3(t−2)	57%	0.439%	Imf 11(t−1)	13%	0.053%
Imf 3(t−3)	7%	0.150%	Imf 11(t−2)	3%	0.000%
Imf 4(t−1)	57%	0.382%	Imf 11(t−3)	0%	0.000%
Imf 4(t−2)	30%	0.207%	Imf 12(t−1)	10%	0.000%
Imf 4(t−3)	20%	0.360%	Imf 12(t−2)	3%	0.000%
Imf 5(t−1)	7%	0.109%	Imf 12(t−3)	13%	0.000%
Imf 5(t−2)	10%	0.000%	Res(t−1)	17%	0.031%
Imf 5(t−3)	13%	0.112%	Res(t−2)	0%	0.000%
Imf 6(t−1)	3%	0.020%	Res(t−3)	17%	0.000%
Imf 6(t−2)	7%	0.001%			
Imf 6(t−3)	4%	0.003%			

Table S9. Input variables for all tested EEMD-LGP models.

Model	Input Variables
LGP 8.1	P(t), P(t−1), Q(t−1), Q(t−2), Q(t−3), **imf set 1*
LGP 8.2	P(t), Q(t−1), Q(t−2), **imf set 1*
LGP 8.3	P(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), **imf set 1*
LGP 8.4	P(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), Q(t−6), **imf set 1*
LGP 8.5	P(t), Q(t−1), Q(t−2), Q(t−3), **imf set 3*
LGP 8.6	P(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), **imf set 3*
LGP 8.7	P(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), **imf set 3*
LGP 8.8	P(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), **imf set 4*
LGP 8.9	P(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), **imf set 2*
**imf set 1*	IMF 1(t−1), IMF 1(t−2), IMF 1(t−3), IMF 2(t−1), IMF 2(t−2), IMF 2(t−3), IMF 3(t−1), IMF 3(t−2), IMF 3(t−3), IMF 4(t−1), IMF 4(t−2), IMF 4(t−3), IMF 5(t−1), IMF 5(t−2), IMF 5(t−3), IMF 6(t−1), IMF 6(t−2), IMF 6(t−3), IMF 7(t−1), IMF 7(t−2), IMF 7(t−3), IMF 8(t−1), IMF 8(t−2), IMF 8(t−3), IMF 9(t−1), IMF 9(t−2), IMF 9(t−3), IMF 10(t−1), IMF 10(t−2), IMF 10(t−3), IMF 11(t−1), IMF 11(t−2), IMF 11(t−3), IMF 12(t−1), IMF 12(t−2), IMF 12(t−3), Residue(t−1), Residue(t−2), Residue(t−3)
**imf set 2*	**imf set 1*, but only IMF 1(t−1) *t*o IMF 4(t−3)
**imf set 3*	**imf set 1*, but only IMF 1(t−1) *t*o IMF 5(t−3)
**imf set 4*	**imf set 1*, but only IMF 1(t−1) *t*o IMF 6(t−3)

Table S10. Statistical parameters for EEMD-LGP models depicted in Table S9.

Model		AARE	*E*	*R*	NMBE	NRMSE	%MF	E_{per}
LGP 8.1	Training	15.909	0.982	0.991	0.763	0.133	−1.674	0.893
	Validation	17.463	0.979	0.989	1.139	0.234	−0.558	0.886
LGP 8.2	Training	15.295	0.980	0.990	−1.064	0.230	−4.510	0.882
	Validation	16.277	0.978	0.989	−0.985	0.239	−6.825	0.881
LGP 8.3	Training	14.483	0.981	0.991	0.476	0.227	−4.868	0.885
	Validation	14.566	0.980	0.990	0.819	0.227	−6.702	0.883
LGP 8.4	Training	19.440	0.981	0.990	0.037	0.227	−3.443	0.885
	Validation	21.152	0.976	0.988	0.952	0.249	−9.031	0.871
LGP 8.5	Training	13.454	0.984	0.992	−0.987	0.207	−1.500	0.905
	Validation	14.232	0.981	0.991	−0.449	0.220	−7.653	0.899
LGP 8.6	Training	16.877	0.980	0.990	0.664	0.230	−4.938	0.883
	Validation	18.348	0.976	0.988	1.273	0.247	−9.295	0.873
LGP 8.7	Training	12.800	0.982	0.991	0.225	0.219	−7.769	0.893
	Validation	13.570	0.979	0.989	0.590	0.234	−8.534	0.887
LGP 8.8	Training	14.316	0.982	0.991	0.589	0.219	−7.302	0.894
	Validation	15.290	0.978	0.989	0.886	0.235	−9.209	0.886
LGP 8.9	Training	19.259	0.982	0.991	1.200	0.220	−3.755	0.893
	Validation	20.576	0.977	0.988	1.818	0.240	−13.072	0.880

LGP-10

Table S11. Output and input variables for LGP as standalone application for predicting ahead.

Model	Time-Step	Input Variable
LGP 10.1	Q(t+1)	P(t), P(t−1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3)
	Q(t+2)	P(t), P(t−1), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3)
	Q(t+3)	P(t), P(t−1), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3)

Table S12. Statistical parameters for LGP standalone models for predicting ahead as depicted in Table S11.

Model	Time-Step		AARE	E	R	NMBE	NRMSE	%MF	E_{per}
LGP 10.1	Q(t+1)	Training	40.293	0.811	0.901	−3.463	0.706	−0.0450	0.277
		Validation	40.921	0.814	0.902	−0.205	0.689	12.788	0.371
	Q(t+2)	Training	64.614	0.631	0.794	0.415	0.989	−4.823	0.084
		Validation	63.781	0.619	0.787	0.533	0.986	0.321	0.115
	Q(t+3)	Training	115.428	0.424	0.651	1.751	1.236	−9.053	0.032
		Validation	112.681	0.421	0.649	2.206	1.215	−15.141	0.055

Table S13. Output and input variables for LGP as standalone with updated precipitation record to account for delays between actual and predicted streamflow (forecasted rainfall would need to be used).

Model	Time-Step	Input Variables
LGP 10.2	Q(t+1)	P(t+1), Q_{pred}(t), Q(t−1)
	Q(t+2)	P(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1)
	Q(t+3)	P(t+3), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1)
	Q(t+4)	P(t+4), Q_{pred}(t+3), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1)
	Q(t+5)	P(t+5), Q_{pred}(t+4), Q_{pred}(t+3), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1)
	Q(t+6)	P(t+6), Q_{pred}(t+5), Q_{pred}(t+4), Q_{pred}(t+3), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1)

Table S14. Statistical parameters for models described in Table S13 which used LGP as a standalone application to predict streamflow multiple time-steps ahead.

Model	Time-Step		AARE	E	R	NMBE	NRMSE	%MF	E_{per}
LGP 10.2	Q(t+1)	Training	41.603	0.849	0.921	−1.127	0.633	−5.617	0.428
		Validation	41.553	0.830	0.911	−2.624	0.659	−0.737	0.432
	Q(t+2)	Training	57.834	0.801	0.895	−0.307	0.727	3.985	0.359
		Validation	60.346	0.754	0.868	−2.900	0.792	17.662	0.329
	Q(t+3)	Training	64.342	0.779	0.883	−4.433	0.765	14.674	0.406
		Validation	62.845	0.724	0.854	−8.633	0.839	62.446	0.348
	Q(t+4)	Training	74.116	0.772	0.879	1.925	0.778	18.85	0.317
		Validation	76.987	0.742	0.861	−1.423	0.811	24.972	0.290
	Q(t+5)	Training	88.018	0.769	0.877	1.097	0.784	11.981	0.325
		Validation	93.604	0.736	0.860	−1.833	0.821	37.554	0.253
	Q(t+6)	Training	99.644	0.771	0.878	1.918	0.780	9.356	0.325
		Validation	104.973	0.739	0.859	-0.593	0.816	44.896	0.286

Table S15. Output and input variables EEMD-LGP models predicting ahead.

Model	Time-Step	
LGP 10.3	Q(t+1)	P(t), $Q_{pred.}$(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
	Q(t+2)	P(t), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
	Q(t+3)	P(t), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
	Q(t+4)	P(t), Q_{pred}(t+3), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
*$imfs$	IMF 1(t−1), IMF 1(t−2), IMF 1(t−3), IMF 2(t−1), IMF 2(t−2), IMF 2(t−3), IMF 3(t−1), IMF 3(t−2), IMF 3(t−3), IMF 4(t−1), IMF 4(t−2), IMF 4(t−3), IMF 5(t−1), IMF 5(t−2), IMF 5(t−3)	

Table S16. Statistical parameters for models described in Table S15 which used a EEMD-LGP design and the known precipitation record to predict streamflow multiple time-steps ahead.

Model	Time-Step		AARE	E	R	NMBE	NRMSE	%MF	E_{per}
LGP 10.3	Q(t+1)	Training	23.993	0.964	0.982	−0.682	0.308	−10.986	0.780
		Validation	24.045	0.958	0.979	0.581	0.329	−15.684	0.771
	Q(t+2)	Training	33.696	0.927	0.963	−1.432	0.439	−12.001	0.601
		Validation	33.027	0.923	0.961	−0.269	0.442	−9.817	0.639
	Q(t+3)	Training	48.708	0.901	0.949	−1.185	0.512	−11.448	0.539
		Validation	47.832	0.889	0.943	0.333	0.531	−15.804	0.522
	Q(t+4)	Training	54.059	0.863	0.929	−1.572	0.604	−14.151	0.412
		Validation	53.799	0.838	0.915	0.587	0.643	−21.699	0.368

Table S17. Output and input variables for EEMD-LGP models that would require forecasted precipitation.

Model	Time-Step	
LGP 10.4	Q(t+1)	P(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
	Q(t+2)	P(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
	Q(t+3)	P(t+3), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
	Q(t+4)	P(t+4), Q_{pred}(t+3), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
	Q(t+5)	P(t+5), Q_{pred}(t+4), Q_{pred}(t+3), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
	Q(t+6)	P(t+6), Q_{pred}(t+5), Q_{pred}(t+4), Q_{pred}(t+3), Q_{pred}(t+2), Q_{pred}(t+1), Q_{pred}(t), Q(t−1), Q(t−2), Q(t−3), Q(t−4), Q(t−5), *$imfs$
*$imfs$	IMF 1(t−1), IMF 1(t−2), IMF 1(t−3), IMF 2(t−1), IMF 2(t−2), IMF 2(t−3), IMF 3(t−1), IMF 3(t−2), IMF 3(t−3), IMF 4(t−1), IMF 4(t−2), IMF 4(t−3), IMF 5(t−1), IMF 5(t−2), IMF 5(t−3)	

Table S18. Statistical parameters for models in Table S17 which used a EEMD-LGP design to predict streamflow multiple time steaps ahead while requiring forecasted precipitation data.

Model		Time-Step	AARE	*E*	*R*	NMBE	NRMSE	%MF	E_{per}
LGP 10.4	Q(*t*+1)	Training	25.575	0.963	0.981	−0.214	0.312	−12.988	0.774
		Validation	25.715	0.954	0.977	0.104	0.342	−11.459	0.752
	Q(*t*+2)	Training	30.872	0.917	0.958	−3.112	0.469	−14.582	0.543
		Validation	30.092	0.911	0.954	−2.641	0.560	−2.806	0.516
	Q(*t*+3)	Training	54.125	0.896	0.947	−3.366	0.526	−7.299	0.548
		Validation	54.363	0.877	0.936	−2.641	0.561	−2.806	0.516
	Q(*t*+4)	Training	58.362	0.876	0.936	−6.096	0.574	−9.468	0.474
		Validation	59.621	0.845	0.919	−5.067	0.629	7.249	0.408
	Q(*t*+5)	Training	63.922	0.859	0.927	−6.310	0.612	−4.615	0.460
		Validation	66.684	0.816	0.903	−5.462	0.684	−6.478	0.367
	Q(*t*+6)	Training	51.661	0.850	0.922	−1.927	0.633	6.473	0.391
		Validation	51.945	0.811	0.901	−0.664	0.695	−1.090	0.322