

Supporting Information for “Regionalization in a global hydrologic deep learning model: from physical descriptors to random vectors”

Xiang Li¹, Ankush Khandelwal², Xiaowei Jia³, Kelly Cutler², Rahul Ghosh²,

Arvind Renganathan², Shaoming Xu², Kshitij Tayal², John Nieber¹,

Christopher Duffy⁴, Michael Steinbach², Vipin Kumar²

¹Department of Bioproducts and Biosystems Engineering, University of Minnesota Twin Cities, St.Paul, MN, USA

²Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, MN, USA

³School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, USA

⁴Department of Civil and Environmental Engineering, Pennsylvania State University, State College, PA, USA

Contents of this file

1. Tables S1 to S7

Introduction

This supplement information document report the p-value statistics regarding all NSE comparison in the paper — “Regionalization in a global hydrologic deep learning model: from physical descriptors to random vectors”. The captions for each p-value table below marks the corresponding NSE comparison table in the paper (from Table 3 to Table 10). All “NA”s mark the corresponding the p-value entries comparing with itself and thus not able to calculate it and reports ”NA”.

Tables S1 to S7**Table S1.** The p-value table for Table 3 and Table 10 in the main article.

Catchment static vectors	p-value (to EP)	p-value (to EG-512)
27-d physical vectors (EP)	NA	9.25e-06
512-d random vectors (EG-512)	9.25e-06	NA
one-hot vectors (EO)	3.14e-10	0.121
27-d physical descriptors + 512-d embedding (PEA)	1.40e-10	0.003
27-d Gaussian vectors + 512-d embedding (REA)	1.41e-09	2.09e-05

Table S2. The p-value table for Table 4 in the main article

k-basin group		10	50	100
	d	p-value to its EP counterpart		
Gaussian vector (EG-d)	2	1.41e-56	4.05e-58	1.99e-59
	8	5.29e-05	6.69e-51	2.10e-43
	16	2.29e-19	7.47e-35	1.72e-21
	32	2.17e-07	6.31e-06	0.003
	64	1.49e-09	1.08e-17	8.84e-27
	128	7.95e-55	2.66e-58	6.13e-42
	256	2.17e-72	9.55e-65	1.29e-53
	512	4.18e-63	8.24e-59	1.64e-52
	1024	9.46e-31	2.28e-33	2.98e-36
one-hot (EO)		6.05e-45	8.24e-40	1.47e-24
27-d physical descriptors (EP)		NA	NA	NA

Table S3. The p-value table for Table 5 in the main article

Number of training years		1	2	5
Gaussian 512-d (EG-512)	to EP	2.69e-09	0.022	1.67e-07
one-hot (EO)	to EP	0.049	1.65e-36	1.42e-18

Table S4. The p-value table for Table 6 in the main article.

models	p-value (to CP)
CP	NA
CO	9.72e-05
CG-16	0.003
CG-32	2.37e-06

Table S5. The p-value table for Table 7 in the main article

k-basin group		10	50	100
	d	p-value to its CP counterpart		
Gaussian vector (CG-d)	2	2.25e-48	1.20e-56	1.05e-54
	8	4.06e-48	3.34e-30	1.02e-24
	16	6.64e-36	2.13e-14	2.29e-08
	32	1.07e-16	1.01e-05	0.231
	64	7.970e-18	0.0112	0.956
	128	9.72e-40	1.89e-14	0.027
	256	1.78e-64	2.57e-51	1.11e-29
	512	9.23e-73	6.91e-79	7.02e-72
	1024	1.68e-78	2.68e-83	5.50e-82
one-hot (CO)		3.33e-32	3.93e-05	0.173
27-d physical descriptors (CP)		NA	NA	NA

Table S6. The p-value table for Table 8 in the main article

catchment static vectors	p-value (to EP)
(0-d) Without static features	1.09e-75
512-d Gaussian vectors (EG-512)	9.25e-06
9-d climate features	4.21e-72
10-d geology features	7.04e-62
8-d geomorphology features	1.87e-62
27-d physical descriptors (EP)	NA

Table S7. The p-value table for Table 9 in the main article

catchment static vectors	d	p-value (to EP)	p-value (to EG-d, the same d)
mixed Gaussian vector (EM-d)	32	1.55e-08	6.38e-03
	64	2.42e-04	4.75e-17
	128	2.21e-08	3.47e-06
	256	2.44e-08	2.43e-04
	512	5.63e-14	2.81e-07
	1024	3.31e-04	7.19e-11