

## SUPPLEMENTAL MATERIALS

*ASCE Journal of Water Resources Planning and Management*

# Exploring Equity Challenges within Deeply Uncertain Water Supply Investment Pathways in the Federal District of Brazil

Bruna M. Araujo, David F. Gold, Lillian B. Lau, Patrick M. Reed,  
and Conceição M. A. Alves

**DOI:** 10.1061/JWRMD5.WRENG-6353

© ASCE 2024

[www.ascelibrary.org](http://www.ascelibrary.org)

## 1 **Appendix S1. FDB Model: Mass Balance, Risks of Failure, and Performance Objectives**

2

3 The FDB model design is based upon WaterPaths framework, initially applied to the Sedento  
4 Valley illustrative case (Trindade et al. 2020), and adapted for the FDB context by Giacomazzo (2020).  
5 Thus, the FDB model explicitly applies DU Pathways framework, incorporating rule systems that respond  
6 to observed system states. This is provided through a mass-balance model, solved for all water sources in  
7 each service area (summarized in Table S1) and mathematically represented by Eq. (S1):

8

$$x_s^{w+1} = x_s^w + NI^w + SE^w + URO^w - ER^w \cdot AR(x_s^w) - EO^w - S^w - RD^w \quad (S1)$$

9

10 where  $x_s^{w+1}$  is the volume of water stored in the reservoir at the week after the current week  $w$ ,  $NI$  is the  
11 natural inflow into the reservoir from all its tributaries,  $SE$  is a treated sewage effluent discharge directly  
12 or indirectly in the reservoir,  $URO$  is the upstream reservoir total outflow,  $ER$  is a non-dimensional  
13 evaporation rate,  $RA$  is the reservoir area as a function of stored volume,  $EO$  is the environmental  
14 outflow,  $RD$  is the total municipal demand drawn from that reservoir to the modelled service area, and  $S$   
15 is the reservoir spillage, which is set to zero unless the reservoir is completely full. Eq. (S1) is solved on a  
16 weekly basis during all simulation time horizon (40 years).

17 Table S1 also summarizes the main features of the two services areas, Descoberto and Santa  
18 Maria, a division proposed by this work based upon FDB water supply infrastructure system. It is  
19 comprised of five main water production subsystems, divided according to water sources location and  
20 associated infrastructures (catchments, treatment plants and water networks). Descoberto and Santa Maria  
21 are the main subsystems, and together they provide water for over 80% of FDB population. As  
22 demonstrated in Fig. 1 of the main text, these two subsystems present wide socioeconomic disparities  
23 between the population they attend, leading to the definition of two service areas of study that spatially  
24 correspond to the water producing subsystems of Descoberto and Santa Maria.

25

26

27 **Table S1.** Main water sources' features in the Federal District supply system

Service area	Water Supply Infrastructure	Storage Capacity <sup>a</sup> (hm <sup>3</sup> )	% Storage allocated in water urban supply <sup>a</sup>	Water Treatment Plant (WTP) <sup>a</sup>	Water Production Capacity <sup>a</sup> (L/s)	Population served <sup>b</sup> (inhab)
Descoberto	Descoberto Reservoir	72,3	100	Descoberto Stream WTP	5,791.8	1,678,243
	Paranoa Lake	460,5	8	Lago Norte WTP		
Santa Maria	Santa Maria Reservoir	61,3	100	Brasília WTP	3,011.1	703,095
	Bananal Stream	-	-			
	Torto Stream	-	-			

28 <sup>a</sup>Data from ADASA (2018).29 <sup>b</sup>Data from CODEPLAN (2018).

30

31 Another core component of the FDB WaterPaths model is the decision making through state-aware, risk  
32 monitoring metrics known as risk of failures (ROF). ROFs express the system's (or service area) ability to  
33 meet required water demand in each time basis. The FDB model presents two ROFs: first, the estimated  
34 probability that the service area total storage falls below a critical level in the next  $T_{rof}$  weeks, if  
35 hydrological conditions from the last 50 years occurs. Second, the estimated probability that service area  
36 demand exceeds 90% of its total treatment capacity. Short term ROFs are calculated with  $T_{rof}$  equals to 52  
37 weeks, and triggers educational campaigns, rationing/contingency tariffs or water transfers if  $\theta_{gr}$ ,  $\theta_{\Delta gr}$  or  
38  $\theta_{gt}$  thresholds are crossed, respectively. Long term ROFs are calculated when  $T_{rof}$  corresponds to 78  
39 weeks, and trigger infrastructure construction in case  $\theta_{ci}$  threshold is crossed. ROFs mathematical  
40 computation is based upon the work of Trindade et al. (2020), and expressed in Eqs. (S2)–(S4):

$$x_{rof,j}^w = \frac{1}{N_{rof}} \sum_{y'=0}^{N_{rof}} f_{y',j}^w(NI^{y'}, E^{y'}) \quad (S2)$$

41 where,

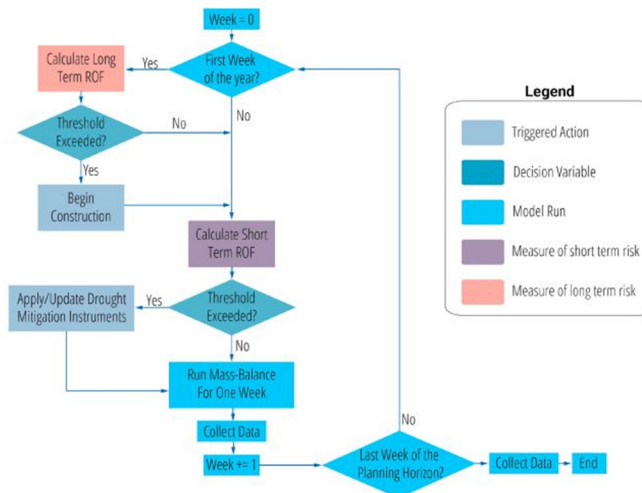
$$f_{y',j}^w = \begin{cases} 0 & \forall w' \in \{(y', w), \dots, (y', w + T_{rof})\}: \frac{x_{s',j}^{y,w'}}{C_j} \geq s_c \\ 1 & \text{otherwise} \end{cases} \quad (S3)$$

42 and,

$$x_{s',j}^{y,w'} = f(C_j, UD_j^w, NI_j^{y',w'}, E_j^{y',w'}, W_j^{y',w'} | \Psi_s) \quad (S4)$$

43 In Eqs. (S2)–(S4),  $w'$  and  $y'$  denote a week and a year simulated with historical data,  $x_{rof,j}^w$   
44 represents ROF for service area  $j$  (Descoberto or Santa Maria) in current week  $w$ ,  $f_{y',j}^w$  is a binary variable  
45 that indicates failure or success during the simulation with data from past year  $y'$ . A failure is identified  
46 when the total water stored in realization  $y'$ ,  $x_{s',j}^{y,w'}$ , divided by storage capacity  $C_j$ , drops below a critical  
47 level. Variable  $x_{s',j}^{y,w'}$  is the vector of storage states determined in one year-long ROF simulations, using  
48 recorded hydrologic data from past year  $y'$ , defined considering storage capacity,  $C_j$ , unrestricted demand,  
49  $UD_j^w$ , recorded natural inflows, evaporation rates and reservoir spillage,  $NI_j^{y',w'}$ ,  $E_j^{y',w'}$  and  $W_j^{y',w'}$ ,  
50 respectively, calculated in year  $y'$  prior to current week  $w$  used in one of the  $N_{rof}$  simulations. Variable  
51  $\Psi_s$  the the vector of DU factors. Fig. S1 presents a schematic view of the ROF calculation and mass  
52 balance loop in WaterPaths.

53



55  
 56 **Fig. S1.** WaterPaths ROF calculation, mass balance loop, and its relationship with decision-making associated with  
 57 short- and long-term management actions. (Reprinted from *Environmental Modelling & Software*, Vol. 132, B. C.  
 58 Trindade, D. F. Gold, P. M. Reed, H. B. Zeff, and G. W. Characklis, “Water pathways: An open source stochastic  
 59 simulation system for integrated water supply portfolio management and infrastructure investment planning,”  
 60 104772, (c) 2020, with permission from Elsevier.)

61

62 As stated in Eq. (3) in the main text, a policy is defined as a set of risk tolerance limits for ROFs  
 63 that, if crossed, trigger associated management measures. While short-term ROF thresholds are related to  
 64 drought mitigation instruments (educational campaigns, rationing and contingency tariffs), long term  
 65 ROF limits can trigger infrastructure construction, whose options for each service area are presented in  
 66 Table S2. The ranges in which the decision variables that compose a candidate policy can vary are  
 67 presented in Tables S3 and S4.

68

69

70 **Table S2.** Candidate supply infrastructure or supply expansion in Federal District water system

Service area	Water supply infrastructure	Description	Cost (10 <sup>6</sup> R\$)	Water Supply(L/s)
Descoberto	Corumba System - 1st phase	New Water Production System	276.50	1,400
	Corumba System - 2nd phase	Water Treatment Expansion	222.10	1,400
	Corumba System - 3rd phase	Water Treatment expansion/new pipelines	251.40	1,200
	Descoberto Reservoir Expansion	1.5 meter spillway level raise	7.50	400
Santa Maria	Paranoa System - 1st phase	New Water Treatment Plant in Paranoá Lake	60.30	700
	Paranoa System - 2nd phase	Water Treatment Expansion	60.30	700
	Paranoa System - 3rd phase	Water Treatment Expansion	60.30	700

71 Sources: Data from GDF (2017); CAESB (2019).  
 72  
 73  
 74  
 75

76 **Table S3.** Lower and upper thresholds for long- and short-term decision variables of the candidate policies

Decision variables	Lower bound	Upper bound
Water consumption restriction trigger for Descoberto - $\theta_{gr}$	0.1%	100%
Water consumption restriction trigger for Santa Maria - $\theta_{gr}$	0.1%	100%
Second stage water consumption restriction trigger for Descoberto - $\theta_{\Delta gr}$	0.1%	100%
Second stage water consumption restriction trigger for Santa Maria - $\theta_{\Delta gr}$	0.1%	100%
Water transfer trigger for Descoberto - $\theta_{gt}$	0.1%	100%
Water transfer trigger for Santa Maria - $\theta_{gt}$	0.1%	100%
Annual reserve fund contribution for Descoberto as percentage of annual revenue - $\theta_{accf}$	0%	10%
Annual reserve fund contribution for Santa Maria as percentage of annual revenue - $\theta_{accf}$	0%	10%
Infrastructure construction long-term ROF trigger for Descoberto - $\theta_{ci}$	0.1%	100%
Infrastructure construction long-term ROF trigger for Santa Maria - $\theta_{ci}$	0.1%	100%

77  
 78  
 79  
 80

81 **Table S4.** Construction order ranges for each infrastructure option

Service area	Decision variables (ICO)	Lower bound	Upper bound
Descoberto	Corumba System - 1 <sup>st</sup> phase	1 <sup>st</sup>	4 <sup>th</sup>
	Corumba System - 2 <sup>nd</sup> phase	1 <sup>st</sup>	4 <sup>th</sup>
	Corumba System - 3 <sup>rd</sup> phase	1 <sup>st</sup>	4 <sup>th</sup>
	Descoberto Reservoir Expansion	1 <sup>st</sup>	4 <sup>th</sup>
Santa Maria	Paranoa System - 1 <sup>st</sup> phase	1 <sup>st</sup>	3 <sup>rd</sup>
	Paranoa System - 2 <sup>nd</sup> phase	1 <sup>st</sup>	3 <sup>rd</sup>
	Paranoa System - 3 <sup>rd</sup> phase	1 <sup>st</sup>	3 <sup>rd</sup>

82

83

84 DU many-objective optimization searches for best performing policies that minimizes function F  
 85 [Eq. (1)], and thus maximizes water supply reliability ( $f_{REL}$ ), minimizes water-use restriction frequency  
 86 ( $f_{RF}$ ), minimizes infrastructure net present value ( $f_{INPV}$ ), minimizes the peak financial cost of drought  
 87 mitigation and debt payments ( $f_{FC}$ ), and minimizes the worst-case cost of drought mitigation actions  
 88 ( $f_{WCC}$ ). Each objective's formulation is presented as follows.

89 Reliability (REL) represents the fraction of states of the world in which reservoir levels drop  
 90 below 20% of its maximum capacity in any given week (failure condition):

$$\max f_{REL} = \min_j \left[ \min_y \left( \frac{1}{N_r} \sum_{i=1}^{N_r} g_{i,j}^y \right) \right] \quad (S5)$$

91 where,

$$g_{i,j}^y = \begin{cases} 0 & \forall w: \frac{x_{s,i,j}^{y,w}}{C_j} \geq s_c \\ 1 & otherwise \end{cases} \quad (S6)$$

92 where  $N_r$  is the number of realizations for one function evaluation and  $g_{i,j}^y$  is a binary function that  
 93 assumes zero value if, in a given year of a specific realization, there was a week when reservoir storage  
 94 fell below  $s_c$  capacity, and 1 otherwise.

95 Restriction frequency (RF) is the fraction of years over the planning horizon (40 years for the  
 96 FDB model) in which at least one week presents use of water restrictions:

$$\min f_{RF} = \max_j \left[ \frac{1}{N_{ys} \cdot N_r} \sum_{i=1}^{N_r} \sum_{y=1}^{N_{ys}} h_{i,j}^y \right] \quad (S7)$$

97 where

$$h_{i,j}^y = \begin{cases} 0 & \forall w: x_{s,i,j}^{y,w} \leq \theta_{rt,j} \\ 1 & \text{otherwise} \end{cases} \quad (S8)$$

98 where  $h_{i,j}^y$  represents the adoption of water use restrictions in a week of a given year of a particular  
 99 realization, and 1 otherwise.

100 Infrastructure net present value (INPV) represents the average net present cost of all  
 101 infrastructures built across all realizations:

$$\min f_{INPV} = \frac{1}{N_r} \sum_{i=1}^{N_r} \sum_{y=1}^{BM} \frac{PMT}{(1+d)^y} \quad (S9)$$

102 where BM is the bond term, d is the discount rate (4%), y is the year of the debt service payment (PMT)  
 103 since the bond was issued, with PMT being calculated by:

$$PMT = \frac{P[BR(1+BR)^{BM}]}{[(1+BR)^{BM} - 1]} \quad (S10)$$

104 where P is the principal, BR is the interest rate and BT is the bond term.

105 The fourth objective, Peak Financial Cost (CF), represents the expected yearly cost of the  
 106 portfolio applied to manage risks over the planning horizon, including revenue losses from restrictions,  
 107 transfer costs, contingency funds contributions and debt repayment:

$$\min f_{CF} = \max_j \left[ \frac{1}{N_{ys} \cdot N_r} \sum_{i=1}^{N_r} \sum_{y=1}^{N_{ys}} SYC_{i,j}^y \right] \quad (S11)$$

108 Next,



$$\begin{aligned}
& SYC_{i,j}^y \\
& = \frac{\sum_{c \in C_j} PMT_{i,j,c} + IP_{i,j}^y + \theta_{acfc,j} \cdot ATR_{i,j}^y + (RL_{i,j}^y + TC_{i,j}^y - IC_{i,j}^y - CF_{i,j}^y, 0)}{ATR_{i,j}^y} \quad (S12)
\end{aligned}$$

109 where  $SYC_{i,j}^y$  correspond to yearly costs for service area j,  $PMT_{i,j,c}$  is the debt payment for infrastructure  
110 option c,  $ATR_{i,j}^y$  is the total annual volumetric revenue,  $IP_{i,j}^y$  is the insurance contract cost in year y,  $RL_{i,j}^y$   
111 is revenue losses due to water restriction,  $TC_{i,j}^y$  is the yearly water transfer costs,  $IC_{i,j}^y$  represents insurance  
112 coverage value, and  $CF_{i,j}^y$  is the value available in contingency fund.

113 The last objective, Worst First Percentile Cost (WCC) represents the 1% highest year drought  
114 management costs across the planning horizon of all SWOs:

$$\min f_{WCC} = \max_j \{ \text{quantile}_{i \in N_r} (SYC_{i,j}, 0.99) \} \quad (S13)$$

115 where

$$SYC_{i,j}^y = \max \frac{RL_{i,j}^y + TC_{i,j}^y - \theta_{acfc,j} \cdot ATR_{i,j}^y - YIPO_{i,j}^y, 0}{ATR_{i,j}^y} \quad (S14)$$

116 where IP is the insurance contract cost in year y, YIPO is the total insurance payout over year y.

117 As already described in Many-Objective Optimization Section, the DU Optimization was  
118 executed using Borg MOEA (Hadka and Reed 2013), a multi-objective evolutionary algorithm that  
119 applies epsilon ( $\epsilon$ )-dominance archiving (Laumanns et al. 2002), stagnation detection, and randomized  
120 restarts to avoid local optima and overcome dominance resistance (Hadka and Reed 2013; Hanne 2001).

121 The significance for  $\epsilon$ -dominance for each objective evaluation is presented in Table S5.

122

123 **Table S5.** The  $\epsilon$ -dominance for objectives' evaluation

Objective	$\epsilon$ -dominance
Reliability	0.001
Restriction Frequency	0.005
Infrastructure Net Present Cost	10.000.000
Peak Financial Cost	0.002
Worst First Percentile Cost	0.005

## 124 **Appendix S2. Building DU SOW Components**

125

126 As already described in Methods Section, DU Optimization and Re-evaluation involves simulating  
127 each candidate policy over a large ensemble of DU SOWs, comprised of one vector of DU factors and one  
128 hydroclimatic realization (Natural Inflow, Natural Evaporation and Water Demand synthetic series). The  
129 DU vector is arranged as a row of the  $\Psi_s$  matrix, and presents one value for each DU of the FDB context.  
130 Each DU element is sampled using Latin Hypercube Sampling (LHS; McKay et al. 1979) within the ranges  
131 presented in Table S6. The boundaries in which DUs can vary were defined based upon technical meetings  
132 with the management and regulation agencies of the FDB water supply system, CAESB and ADASA,  
133 respectively.

134 The hydroclimatic realizations consist of a combination of 1,000 40-year long synthetic  
135 timeseries of natural inflows, natural evaporation, and water demands for each service areas. The inflows  
136 and evaporation timeseries were built using the Modified Gaussian Fraction Noise (mFGN) method  
137 (Kirsch et al. 2013), which bootstraps data from the historical record by accounting for temporal and  
138 spatial correlations to generate timeseries that represent realistic possible future alternatives. It's  
139 important to point out that the lack of large hydrologic historical data is a common obstacle in many  
140 Brazilian studies, and was overcome with the use of data filling and extension methodology presented in  
141 Souza (2022).

142 Fig. S2 shows the flow duration curves of historic and synthetic data for the water sources of the  
143 two service areas (Descoberto and Corumba reservoirs for Descoberto service area, Paranoa and Santa  
144 Maria reservoirs and Bananal/Torto streams for Santa Maria service area). The figures present the  
145 probability of exceedance based upon yearly data of the synthetic and historic flows. Since it has more  
146 extreme low and high values, the synthetic series appear to expand upon historic observations, without a  
147 bias and following the same hydrologic within-year behavior.

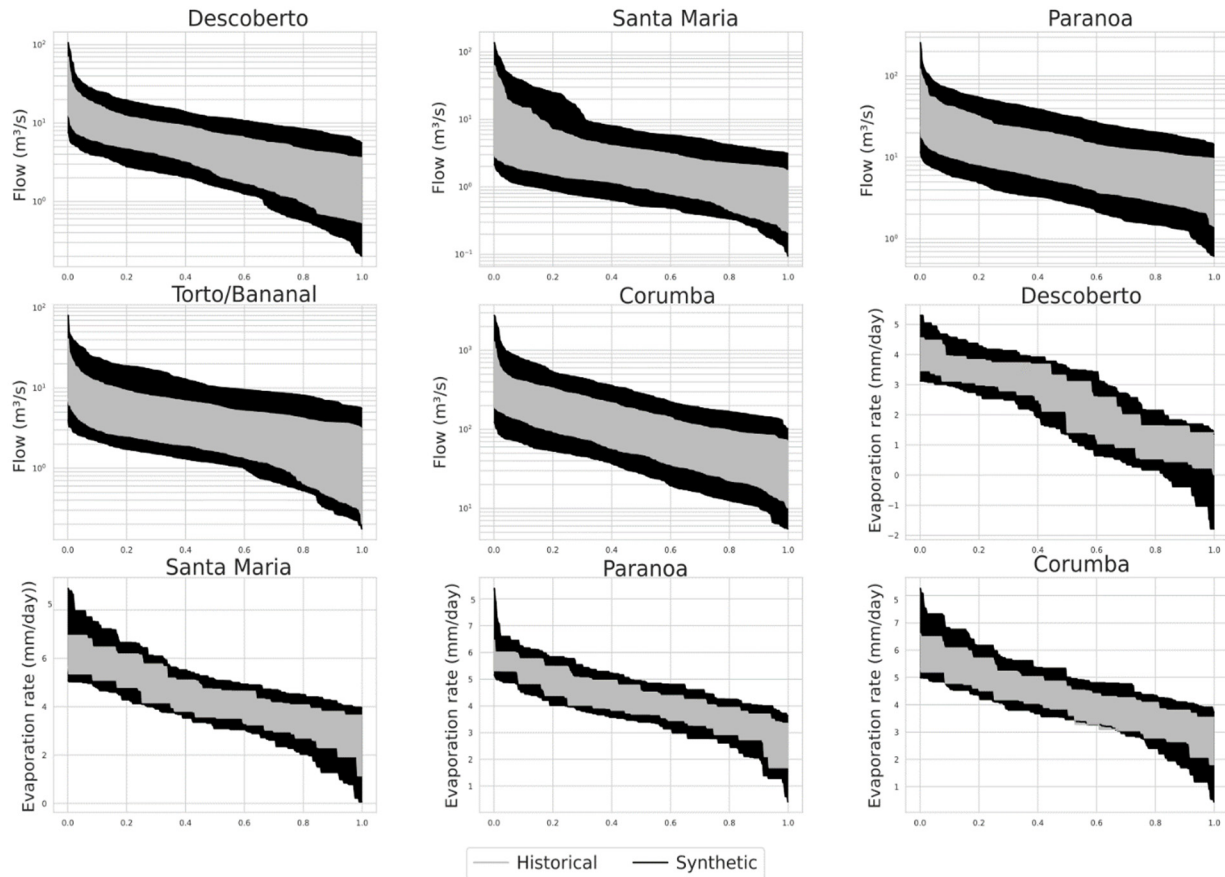
148

149 **Table S6.** Deep uncertainties and respective boundaries defined in a collaborative process with management and  
 150 regulation agencies of the FDB water supply system

Deep uncertainties		Reference Value	Lower Bound		Upper Bound	
			MF <sup>a</sup>	AV <sup>b</sup>	MF	AV
Financial Indexes	Bond interest rate (% py)	12.00	0.84	10.08	1.16	13.92
	Bond term (years)	15.00	0.90	13.50	1.30	19.50
	Discount rate (%)	4.00	0.80	3.20	1.80	7.20
Water and Sanitation Tariffs (R\$/m <sup>3</sup> )	Residential	4.43		4.52		5.14
	Commercial	13.08		13.34		15.17
	Industry	12.02	1.02	12.26	1.16	13.94
	Public services	13.08		13.34		15.17
	Sewage	4.80		4.89		5.57
Drought mitigation instruments effectiveness of water consumption restriction (%)	1st stage	2.00		1.60		2.40
	2nd stage	4.00	0.80	3.20	1.20	4.80
	3rd stage	10.00		8.00		12.00
Time to get Permitting (years)	Corumba - 1st stage	0.00	-	-	-	-
	Corumba - 2nd stage	0.00	-	-	-	-
	Corumba - 3rd stage	5.00	0.60	3.00	1.40	7.00
	Descoberto resevoir expansion	0.00	-	-	-	-
	Paranoa - 1st stage	0.00	-	-	-	-
	Paranoa - 2nd stage	0.00	-	-	-	-
Construction costs (in millions of R\$)	Paranoa - 3rd stage	0.00	-	-	-	-
	Corumba - 1st stage	276.50		276.50		359.45
	Corumba - 2nd stage	222.10		222.10		288.73
	Corumba - 3rd stage	251.40		251.40		326.82
	Descoberto resevoir expansion	7.50	1.00	7.50	1.30	9.75
	Paranoa - 1st stage	60.30		60.30		78.39
Paranoa - 2nd stage	60.30		60.30		78.39	
Paranoa - 3rd stage	60.30		60.30		78.39	

151 Note: <sup>a</sup>MF: Multiplicative Factor - the value effectively applied as boundaries for sampling DU values used in WaterPaths.  
 152 <sup>b</sup>AV: Absolute Value – value obtained when MF multiplies the Reference Value, and represents the actual value of a DU factor  
 153 or boundary.  
 154

155 It's important to emphasize that the aspect of evaporation curves is due to the lack of available  
 156 daily records in the reservoirs' region. Thus, given that the FDB has only two well defined seasons – one  
 157 rainy, hot summer and one dry, cold winter (CODEPLAN 2020) – the assumption of equal daily  
 158 evaporation rates across the months is reasonable for the purposes of this work.

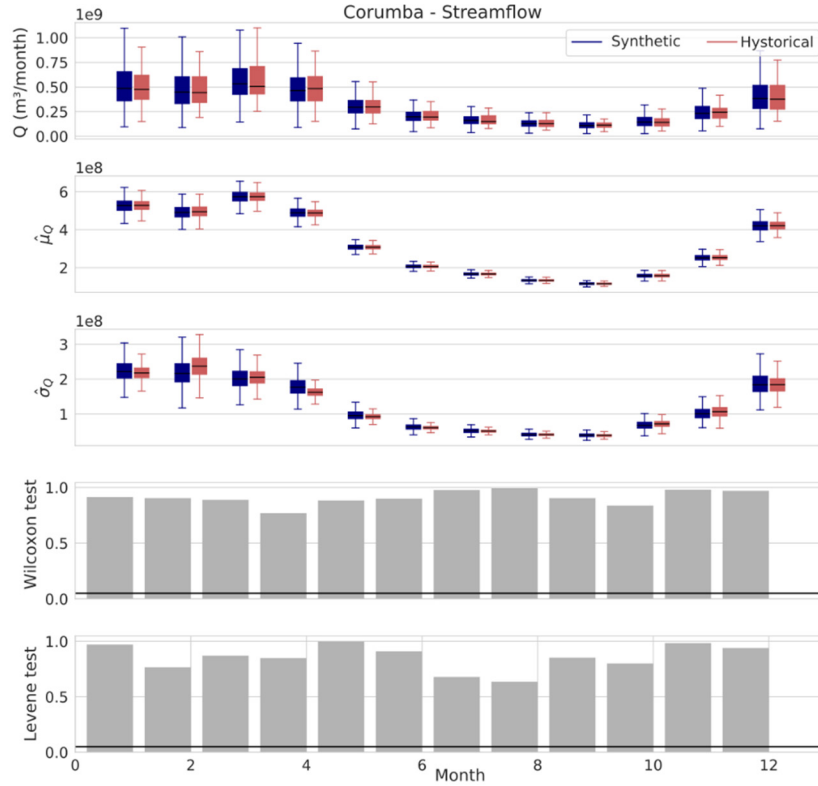


159  
 160 **Fig. S2.** Flow duration curves for natural inflow and natural evaporation synthetic series.

161

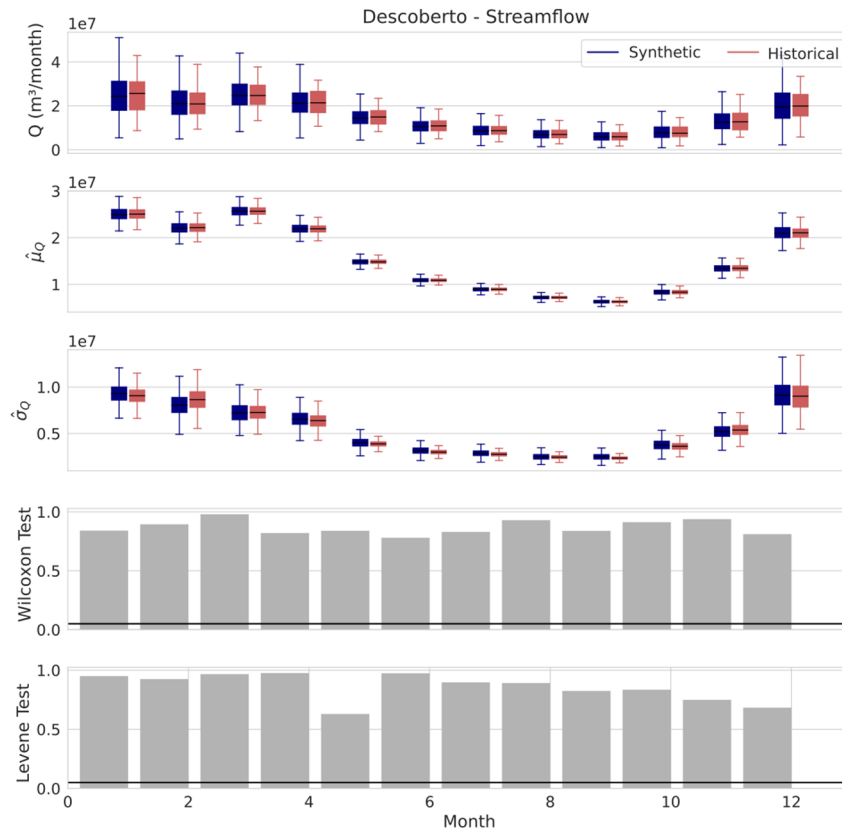
162 Figs. S3–S11 present statistic comparisons of observed and synthetic data, generated from one-  
 163 hundred ensemble of synthetic series with 100-years length, and a bootstrapped historical data with the  
 164 same size and length. We can denote that, for all water sources, streamflow and evaporation synthetic  
 165 series present reasonably close behavior if compared to historical data.

166 Given the non-parametric nature of the method applied for synthetic series generation, Wilcoxon  
 167 rank-sum test and Levene’s test were used to test if synthetic monthly medians and variances were  
 168 statistically different from observations. The associated p-values indicate that none of the synthetic  
 169 momentum differ from the historical at significance level of 0.05.



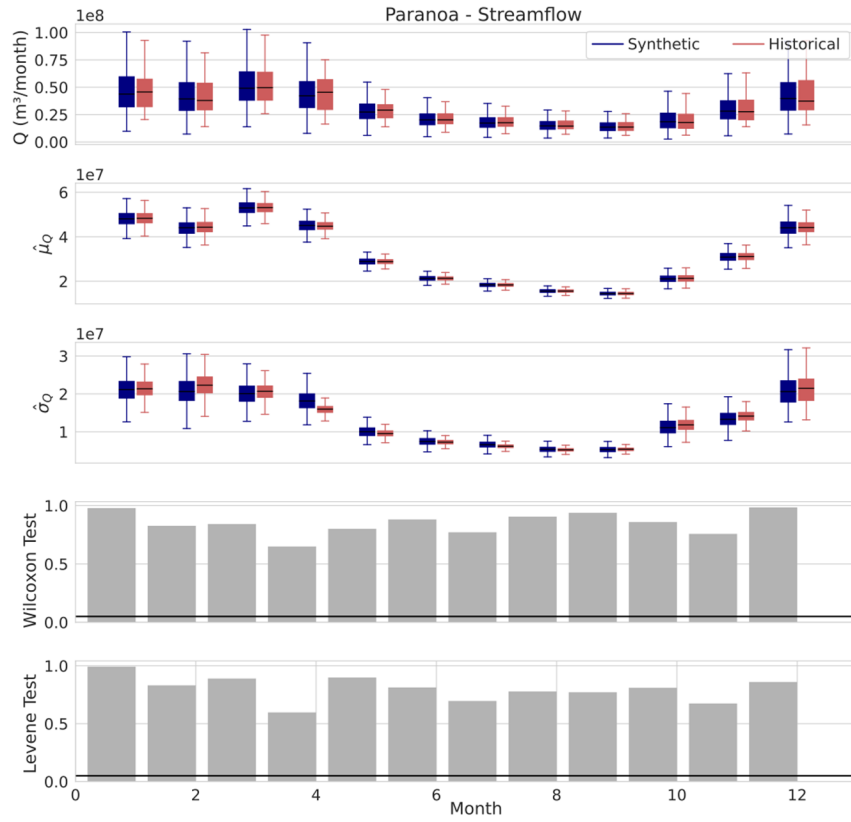
170  
171  
172

Fig. S3. Statistical evaluation for the Corumba synthetic streamflow timeseries.



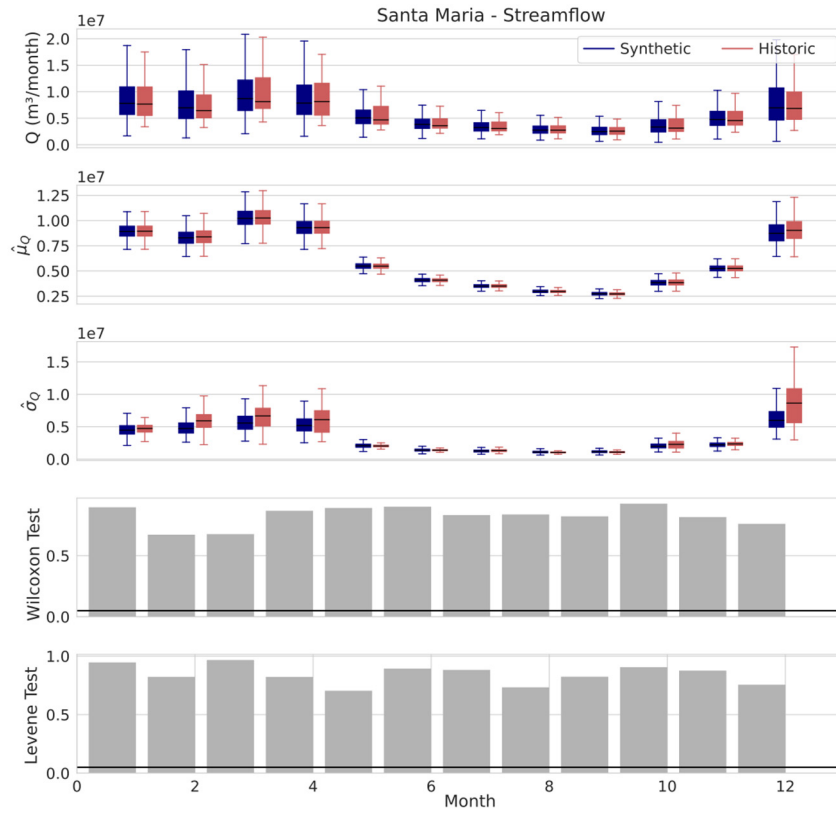
173  
174

Fig. S4. Statistical evaluation for the Descoberto synthetic streamflow timeseries.



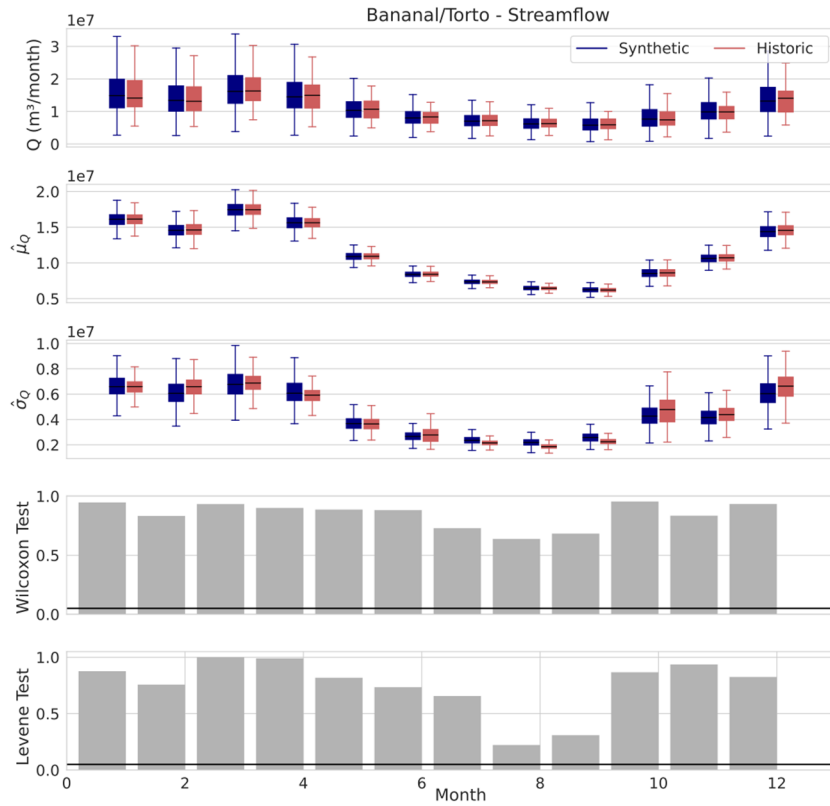
175  
176

Fig. S5. Statistical evaluation for the Paranoa synthetic streamflow timeseries.



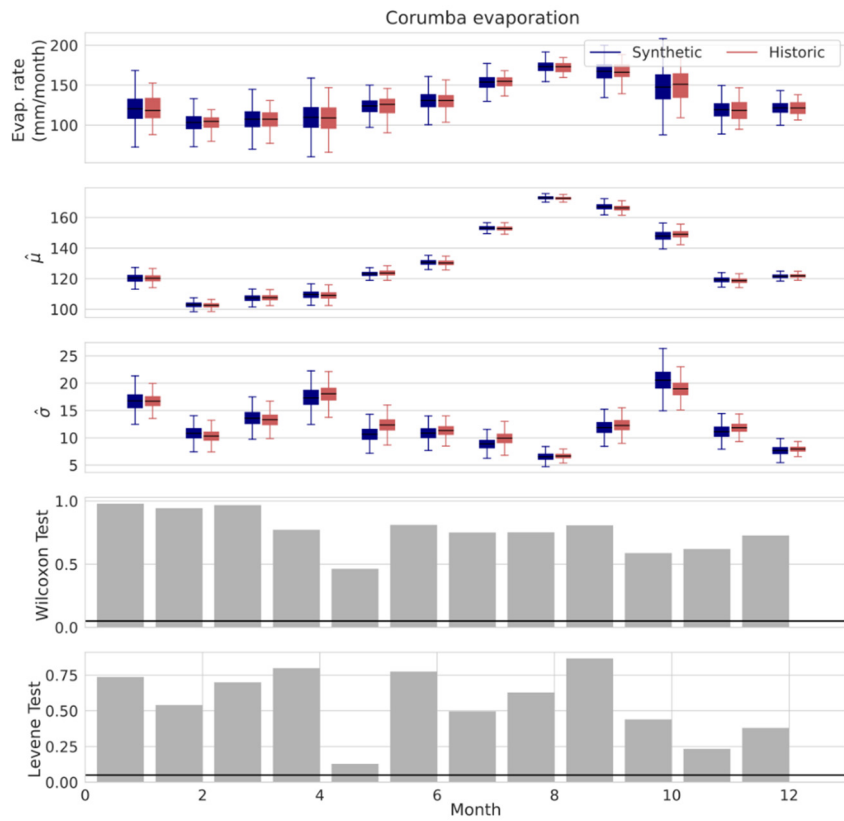
177  
178

Fig. S6. Statistical evaluation for the Santa Maria synthetic streamflow timeseries.



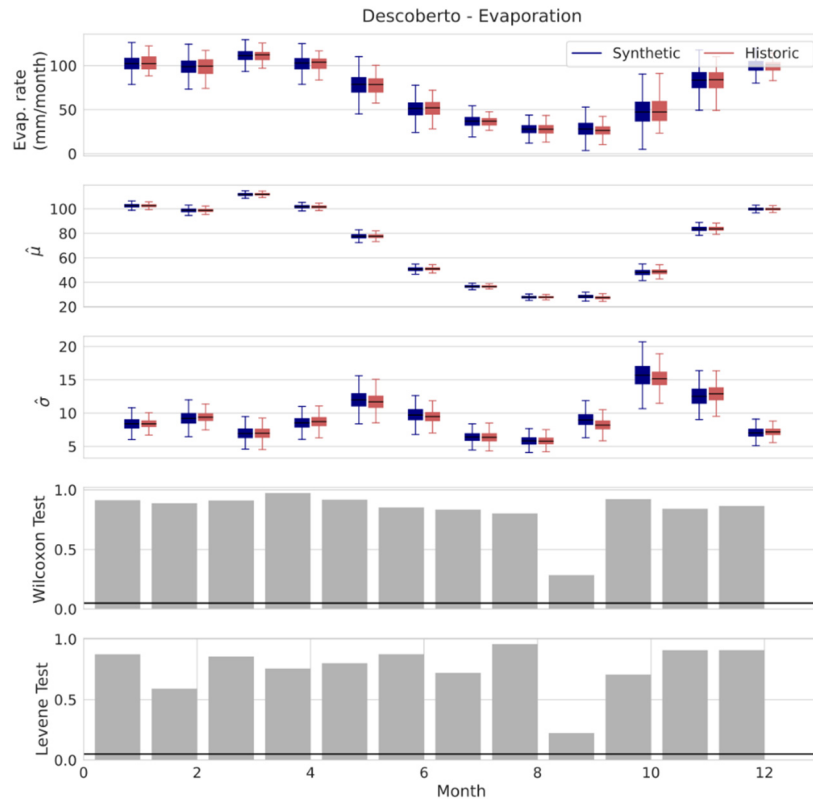
179  
180

Fig. S7. Statistical evaluation for the Bananal-Torto the synthetic streamflow timeseries.



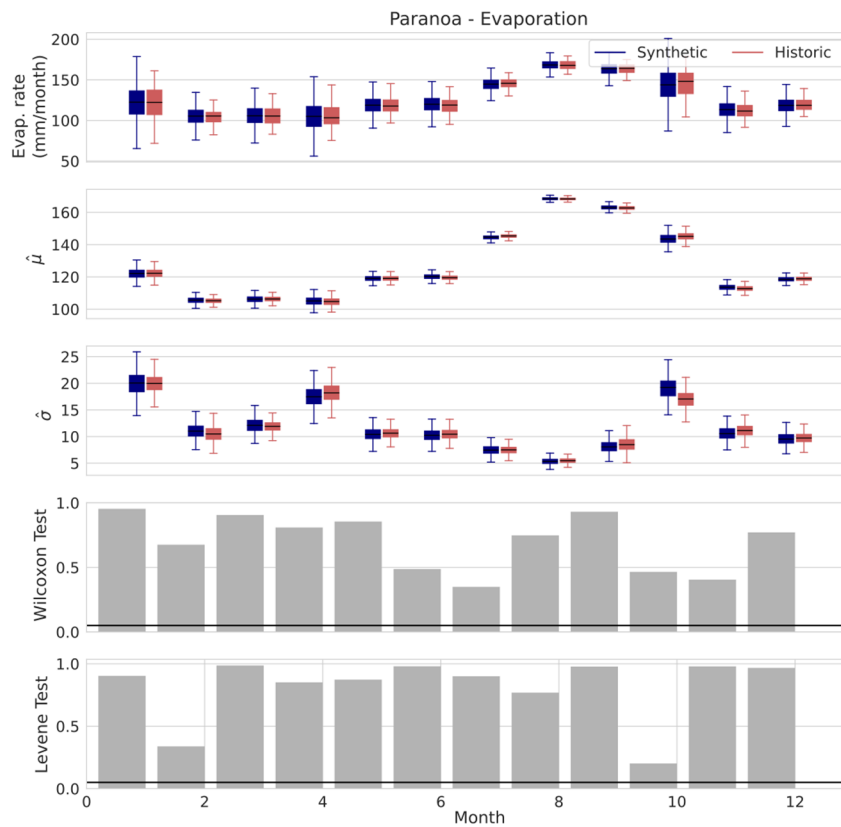
181  
182

Fig. S8. Statistical evaluation for the Corumba synthetic evaporation timeseries.



183  
184

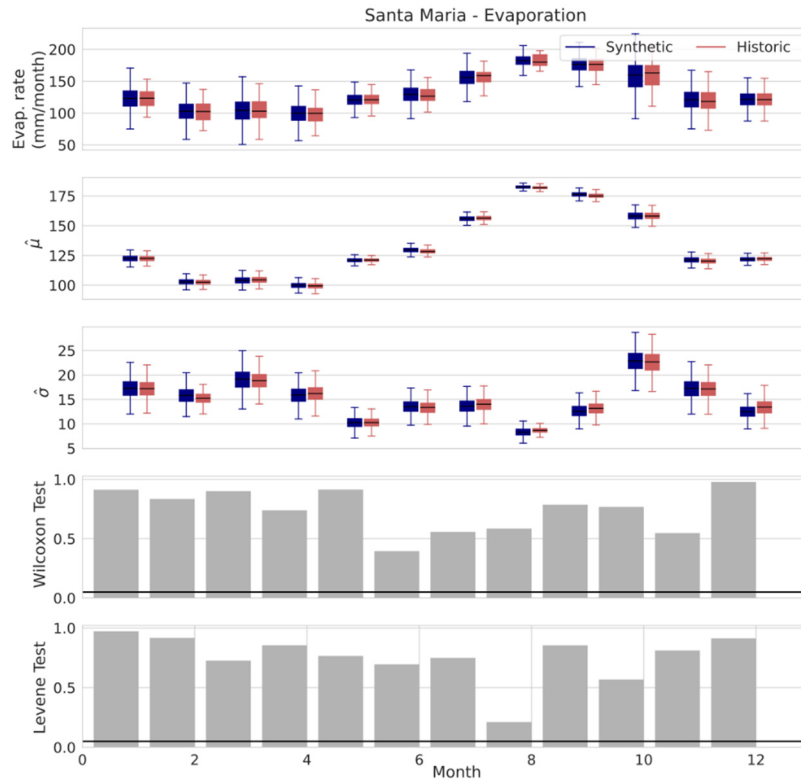
Fig. S9. Statistical evaluation for the Descoberto synthetic evaporation timeseries.



185  
186  
187

Fig. S10. Statistical evaluation for the Paranoa synthetic evaporation timeseries.

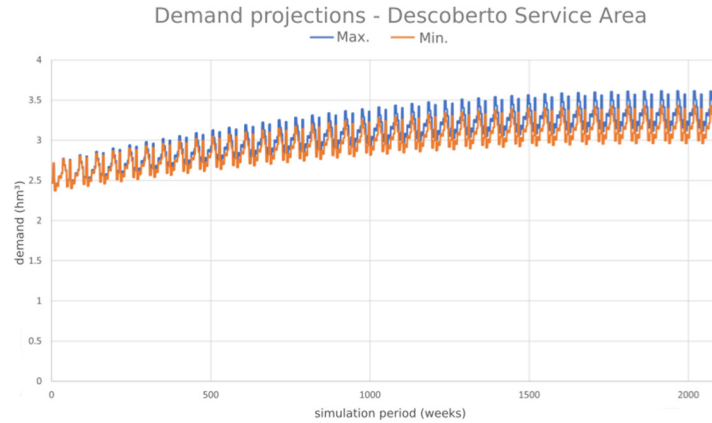




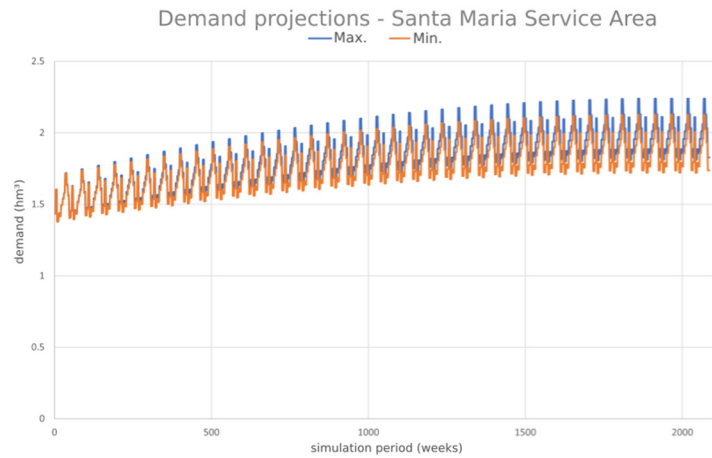
188  
189 **Fig. S11.** Statistical evaluation for the Santa Maria synthetic evaporation timeseries.

190

191 The 1,000 demand time series were built applying different multiplicative factors to a reference  
 192 demand projection. This reference series, in turn, was obtained by taking the Federal District population  
 193 projection for the next 40 years in IBGE, and associating it with demand estimatives extracted from Federal  
 194 District Sanitation Plan of 2017 (GDF 2017). Figures S12 and S13 show the highest and lowest demand  
 195 synthetic series for Descoberto and Santa Maria service area, suggesting reasonable variability given  
 196 population projections for each set of the FDB Administrative Regions.



197  
 198 **Fig. S12.** Upper- and lower-bounds of the demand projections for the Descoberto service area.  
 199  
 200  
 201  
 202



203  
 204 **Fig. S13.** Upper- and lower-bounds of the demand projections overview for the Santa Maria service area.

205  
 206

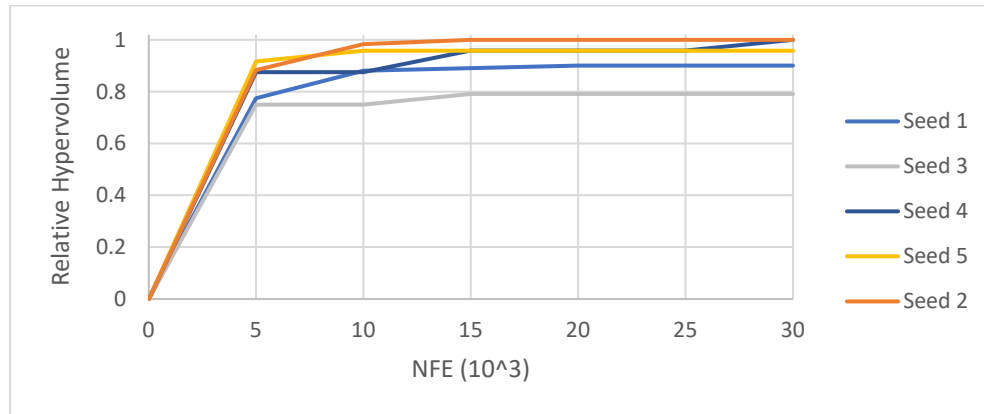
207 **Appendix S3. MOEA Runtime Diagnostics**

208

209 We run multiple instances of the Borg MOEA to ensure the algorithm has overcome any biases  
 210 generated by the random sampling of the initial population (Salazar et al. 2017). In this experiment, we  
 211 ran 5 random seeds using the Master-Worker configuration of the Borg MOEA. The true Pareto set for  
 212 this problem is not known, so we used the relative hypervolume metric to assess the convergence of each  
 213 seed (Zitzler et al. 2003). The relative hypervolume compares the performance of the Pareto sets  
 214 discovered at set checkpoints within each seed to the final “reference set”, which contains non-dominated

215 solutions across all seeds. We conclude that the algorithm has converged to a satisfactory approximation  
216 of the true Pareto set when the hypervolume of each reference set plateaus. Runtime diagnostics for all  
217 seeds are shown in Fig. S14. There was some variance across seeds, but the hypervolume of each  
218 optimization run plateaued after approximately 10,000 function evaluations.

219



220  
221 **Fig. S14.** Runtime diagnostics for optimization.

222  
223  
224

## 225 **Appendix S4. Gradient Boosted Trees Feature Importance**

226

227 We measure feature importance by evaluating how each DU factor reduces leaf impurity during  
228 Gradient Boosted Trees classification. Leaf impurity provides a measure of how “mixed” a leaf of a  
229 classification tree is. A leaf containing all scenarios with the same classification will have a leaf impurity  
230 of 0, and leaf impurity will increase with the increasing fraction of scenarios classified differently. We  
231 calculate the percentage of the total decrease in leaf impurity across all trees due to splits from each DU  
232 factor in the performance of a policy to calculate feature importance. Feature importance for the High  
233 infrastructure and Moderate infrastructure policies are shown in Table S7.

234

235 **Table S7.** Feature importance ranking for the High infrastructure and Moderate Infrastructure compromises

Factor	Feature importance for High Infrastructure compromise	Feature importance for Moderate Infrastructure compromise
Demand Growth	0.432	0.527
Permitting Time	0.001	0.000
Construction Time	0.002	0.000
Bond Term	0.002	0.002
Bond Interest Rate	0.001	0.002
Discount Factor	0.001	0.002
Restriction Effectiveness	0.561	0.467

236  
237

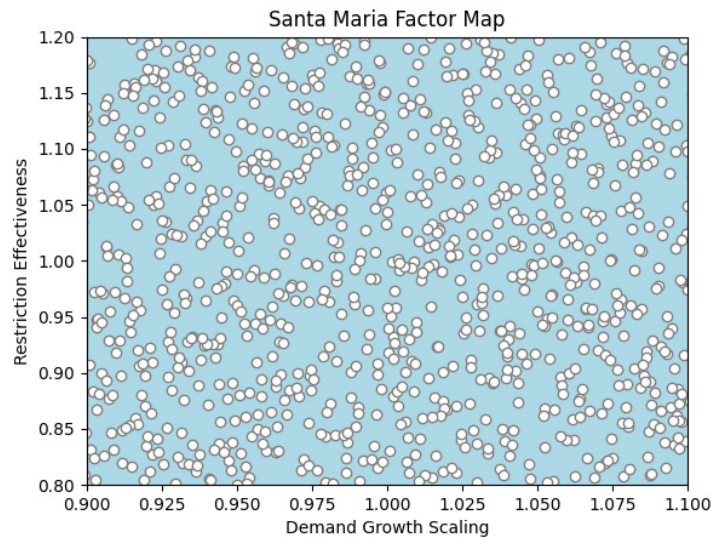
238

239 **Appendix S5. Scenario Discovery for Santa Maria**

240

241 The compromise policies perform extremely well for Santa Maria across all SOWs. Santa Maria does not  
242 fail the robustness criteria (90% reliability, 20% restriction frequency) under any sampled SOW. An factor  
243 map showing Santa Maria’s performance for the moderate infrastructure compromise is shown in Fig. S15.

244



245

246 **Fig. S15.** Factor map for Santa Maria. Each point represents a sampled SOW, with white points representing  
247 SOWs where the policy meets performance criteria, and red points representing SOWs where the policy  
248 fails to meet performance criteria. The policy meets performance criteria under all sampled SOWs.

249 **References**

- 250 ADASA (Regulation Agency for Water, Sewage and Sanitation for the Federal District) [In Portuguese]. CAESB  
 251 (Environmental Sanitation Company of The Federal District) [In Portuguese]. EMATER-DF (Federal District  
 252 Enterprise for Technical Assistance and Rural Extension) [In Portuguese]. SEAGRI (The State Secretariat for  
 253 Agriculture, Supply and Rural Development) [In Portuguese]. 2018. *2016-2018 Water Crisis Management:  
 254 Federal District Experiences* [In Portuguese]. Brasília: ADASA.
- 255 CAESB (Enviromental Sanitation Company of the Federal District [In Portuguese]. 2019. *Joint Technical Note  
 256 number 15/2019* [In Portuguese]. Brasília, Federal District, Brazil.
- 257 CODEPLAN (Federal District Planning Company) [In Portuguese]. 2020. *Federal District Atlas* [In Portuguese].  
 258 Brasília, Federal District, Brazil.
- 259 CODEPLAN (Planning Company of the Federal District). 2018. [In Portuguese]. “District Research by Household  
 260 Sampling” [In Portuguese]. Accessed June 27, 2023. <https://www.codeplan.df.gov.br/pdad-2018>.
- 261 GDF (Federal District Government) [In Portuguese]. 2017. *District Plan for Sanitation and Integrated Management  
 262 of Solid Waste*. Brasília, Federal District, Brazil.
- 263 Giacomazzo, A. P. 2020. Planning and Management Portifolios Assessments of Federal District of Brazil Water  
 264 Supply System in Contexts of Deep Uncertainty [In Portuguese]. Master Degree Dissertation, Brasília:  
 265 University of Brasília.
- 266 Hadka, D., and Reed, P. 2013. “Borg: An auto-adaptive many-objective evolutionary computing framework.” *Evol.  
 267 Comput.*, 21(2), 231–259. [https://doi:10.1162/EVCO\\_a\\_00075](https://doi:10.1162/EVCO_a_00075).
- 268 Hanne, T. 2001. “Global Multiobjective Optimization with Evolutionary Algorithms: Selection Mechanisms and  
 269 Mutation Control.” In EMO 2001, edited by E. Zitzler, L. Thiele, K. Deb, C.A. Coello, D. Corne, Lecture Notes  
 270 in Computer Science, 1993. Berlin: Springer.
- 271 IBGE. (Brazilian Institute of Geography and Statistics) [In Portuguese]. 2020. “Federal District: Population” [In  
 272 Portuguese]. Accessed June 20, 2021. <https://cidades.ibge.gov.br/brasil/df/panorama>.
- 273 Kirsch, B. R. Characklis, G. W. Zeff, H. B. 2013. "Evaluating the Impact of Alternative Hydro-Climate Scenarios  
 274 on Transfer Agreements: Practical Improvement for Generating Synthetic Streamflows". *J. Water Resour. Plan.  
 275 Manag.*, 139 (4), 396–406. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000287](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000287).
- 276 Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. 2002. “Combining convergence and diversity in evolutionary  
 277 multiobjective optimization”. *Evol. Comput.*, 10(3), 263–282. <https://doi.org/10.1162/106365602760234108>
- 278 Mckay, M. D., Beckman, R. J., and Conover, W. J. (1979). “A comparison of three methods for selecting values of  
 279 input variables in the analysis of output from a computer code”. *Technometrics*, 21, 239-245.  
 280 <https://doi.org/10.2307/1268522>.
- 281 Salazar, J. Z., Reed, P. M., Quinn, J. D., Giuliani, M., and Castelletti, A. 2017. “Balancing exploration, uncertainty  
 282 and computational demands in many objective reservoir optimization.” *Adv. Water Resour.*, 109 , 196–210.  
 283 <https://doi.org/10.1016/j.advwatres.2017.09.014>.
- 284 Souza, S. A. (2022). Uncertainties in testing hydrometeorological trends to decision-making concerning adaptation  
 285 measures. [In Portuguese]. PhD Thesis, Brasília: University of Brasília.
- 286 Trindade, B. C., Gold, D. F., Reed, P. M., Zeff, H. B., and Characklis, G. W. 2020. "Water pathways: An open  
 287 source stochastic simulation system for integrated water supply portfolio management and infrastructure  
 288 investment planning". *Environ. Model. Softw.*, 132. <https://doi.org/10.1016/j.envsoft.2020.104772>.
- 289 Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and Da Fonseca, V. G. 2003. “Performance assessment of  
 290 multiobjective optimizers: An analysis and review.” *IEEE Trans. Evol. Comput.*, 7 (2), 117–132. DOI:  
 291 10.1109/TEVC.2003.810758.