Method	OASIS-Brain	Pediatric Airway					
		nasalspine	choana	epiglottic tip	TVC	subglottis	carina
LightGBM	3.0785±0.2852	**50.5081±10.6371**	**42.3419±4.5087**	40.9147±1.3113	**39.5711±2.9905**	30.1161±2.9691	**17.6331±1.7725**
EBM	3.0804±0.1696	65.6239±10.2506	50.5583±4.9465	78.4882±17.5468	102.7091±17.754	104.2232±18.0343	81.3452±20.4006
NAM	3.1834±0.1849	59.4108±4.5717	47.3771±3.4358	70.8585±14.9689	86.5026±16.0979	86.5572±16.1736	80.8666±18.0153
PlainMLP	3.1895±0.1202	59.3684±8.7473	46.6766±3.8046	41.8936±1.2427	43.6047±4.5201	34.1364±3.4129	18.7757±2.7239
GAMLSS	3.1214±0.3022	114.8559±11.0354	48.5584±4.4919	40.8257±1.6604	46.7981±4.6078	37.6076±3.8126	19.3311±2.6193
NAMLSS	3.1006±0.1757	**53.3489±11.3645**	44.6601±4.8858	**39.7347±2.5717**	45.337±9.4324	38.0034±10.8765	18.8011±3.4943
LA-NAM	**3.0619±0.1679**	97.3163±8.4914	56.1675±7.746	92.5769±17.0042	100.0374±28.6009	100.735±31.2878	99.3231±24.7711
Ours_no_prior	3.0757±0.225	64.7846±7.8943	46.037±3.6327	42.3854±3.5059	41.3512±3.2084	31.5489±3.1855	18.0281±2.6473
Ours_part	3.0928±0.1975	79.7719±10.3202	44.3426±3.8319	40.5662±2.6101	39.9039±2.3953	**29.8343±1.7406**	18.5609±1.4272
Ours_full	**3.0708±0.1877**	57.5528±9.8469	**43.0832±4.6037**	**39.1137±2.3979**	**35.7647±2.0409**	**26.1361±1.6574**	**16.0143±2.1051**

Table 2: Quantitative Evaluation of Normalized Brain Volume Regression (OASIS Brain Dataset) and Cross-Sectional Area Regression (Pediatric Airway Dataset) with respect to Mean Absolute Relative Percent Difference (MARPD, %). We also evaluate with respect to different landmarks. The {TVC, subglottic and carina} landmarks are significant landmarks for airway obstruction analysis. **Bold red values** indicate the best scores across all methods. **Bold black values** indicate the 2nd best scores of all methods. *Ours np* refers to `LucidAtlas` without incorporating prior knowledge. *Ours part* denotes our model trained only on complete data, while *Ours imp* represents using the full dataset for training, including missing values. `LucidAtlas` performs best overall.

Method	OASIS-Brain	Pediatric Airway					
		nasalspine	choana	epiglottic tip	TVC	subglottis	carina
PlainMLP	0.8123±0.2113	1.5753±1.3779	0.7499±0.2747	-0.1549±0.1276	-0.8544±0.1313	-1.2154±0.1666	**-1.3672±0.1269**
GAMLSS	0.8184±0.3642	1.4964±0.0977	0.9985±0.1102	-0.1516±0.1991	-0.743±0.0923	-1.1412±0.0893	-1.3632±0.0917
NAMLSS	0.6827±0.0468	1.2118±0.796	0.7778±0.24	**-0.2146±0.1404**	**-1.016±0.1102**	**-1.4053±0.0551**	-1.2714±0.2415
LA-NAM	0.7909±0.0199	1.0725±0.3471	0.9772±0.1086	0.754±0.0204	0.7278±0.0176	0.7262±0.0208	0.7241±0.0284
Ours_no_prior	**0.6703±0.062**	**0.7129±0.543**	0.667±0.1812	-0.1519±0.0939	-1.0061±0.1554	**-1.387±0.209**	-1.366±0.1903
Ours_part	0.6973±0.0623	**0.8338±0.3855**	**0.6482±0.1836**	-0.1513±0.2351	-0.9429±0.1395	-1.3572±0.1222	-1.2978±0.0566
Ours_full	**0.6795±0.0451**	0.8345±0.9108	**0.5959±0.2575**	**-0.2266±0.1232**	**-1.2297±0.141**	-1.0654±1.2994	**-1.3889±0.3671**

Table 3: Quantitative Comparison of Different Ways of Marginalization. NLL is computed between the marginalized covariate interpretation and the data distribution. A ✓in the **Corr.** column indicates that covariate dependence is considered, while ✗signifies that it is ignored. Accounting for covariate dependence improves alignment between covariate interpretation and the data distribution.

Pediatric Airway							
Feat	**Corr.**	**nasalspine**	**choana**	**epiglottic$_t ip$**	**TVC**	**subglottis**	**carina**
AGE	✗	1.0335±0.21	0.6739±0.2276	-0.1435±0.1208	-0.6305±0.1976	-0.8576±0.209	-0.8948±0.1024
AGE	✓	**0.8571±0.8581**	**0.5498±0.2659**	**-0.2535±0.0749**	**-1.1774±0.1022**	**-1.5772±0.1304**	**-1.4041±0.1547**
HEIGHT	✗	1.0912±0.5217	0.7085±0.2158	-0.1917±0.1053	-0.7925±0.1851	-1.0231±0.1918	-0.8743±0.1389
HEIGHT	✓	**0.5881±0.6109**	**0.5644±0.2663**	**-0.2732±0.111**	**-1.2172±0.1309**	**-1.6003±0.1452**	**-1.4548±0.1164**
WEIGHT	✗	1.2684±0.5669	0.8323±0.2044	-0.151±0.0913	-0.5594±0.1265	-0.7481±0.1317	-0.7306±0.103
WEIGHT	✓	**0.7577±0.6689**	**0.6124±0.2241**	**-0.2253±0.0976**	**-1.1127±0.1567**	**-1.4818±0.1704**	**-1.3356±0.0666**

Table 4: Quantitative Comparison of Different Ways of Marginalization. NLL is computed between the marginalized covariate interpretation and the data distribution. A ✓in the **Corr.** column indicates that covariate dependence is considered, while ✗signifies that it is ignored. Accounting for covariate dependence improves alignment between covariate interpretation and the data distribution.

Time	**OASIS Brain**	**Pediatric Airway**					
		nasalspine	**choana**	**epiglottic tip**	**TVC**	**subglottis**	**carina**
T0	**1.6147±0.3043**	**42.1939±23.2475**	**38.1825±10.4036**	52.2004±18.928	**42.7377±13.1737**	**27.059±10.8268**	**18.5224±7.9335**
Pop.	3.2932±0.5446	74.0849±24.6666	40.4904±3.1551	**45.0938±9.1101**	43.2648±4.2845	29.6953±2.8212	18.3675±1.7419
Ind.	**1.3633±0.1764**	**42.7979±23.2923**	**37.539±9.9442**	52.3188±19.109	**41.3312±10.2495**	**24.7931±6.6426**	**16.41±4.0892**

Table 5: Mean Absolute Relative Percent Difference (in %) for Individualized Prediction. **T0** in the **Time** column indicates directly using the observation from the initial time point $T0$ to predict at time $T1$. **Pop.** indicates ignoring the observation at $T0$ and directly using the mean population value $f^m(\boldsymbol{c}, x)$ for individualized prediction for $T1$. **Ind.** indicates our individualized prediction approach illustrated in Sec.3.4.2. Individualized prediction provides the best performance for both datasets and for most landmarks.

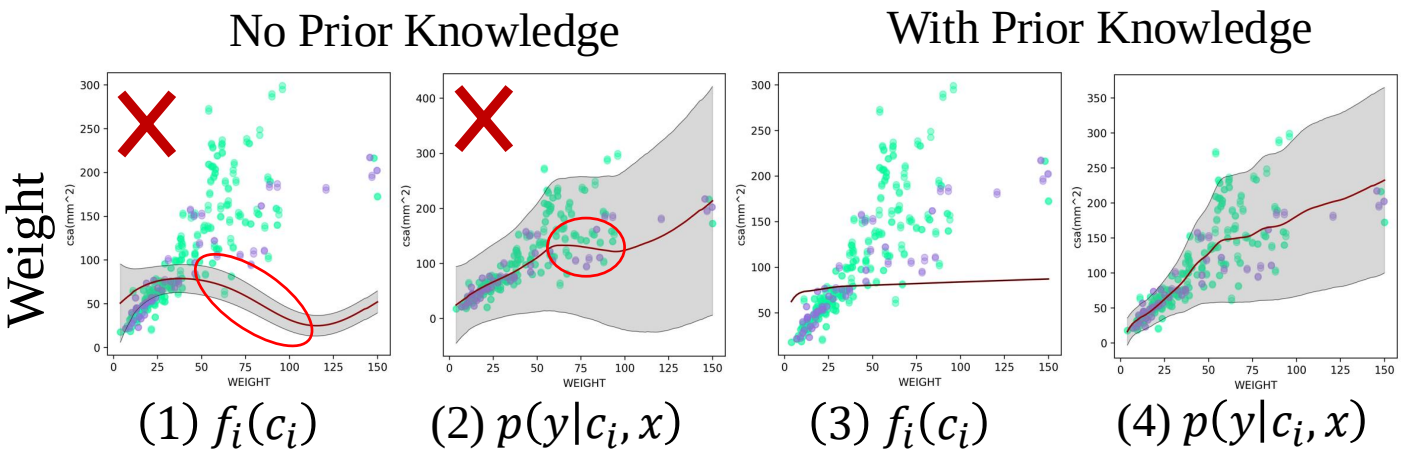

Figure 3: Visualizations of the Effect of Prior Knowledge in `LucidAtlas` at the Subglottis Landmark (Pediatric Airway Dataset). The $\times$ symbol indicates the covariate interpretation contradicts prior knowledge, such as the NAM incorrectly interpreting airway CSA as decreasing with a child's weight. Without incorporating prior knowledge, the model may deviate form our prior assumptions. Without marginalization, to account for covariate dependencies, the data may not be fit well.

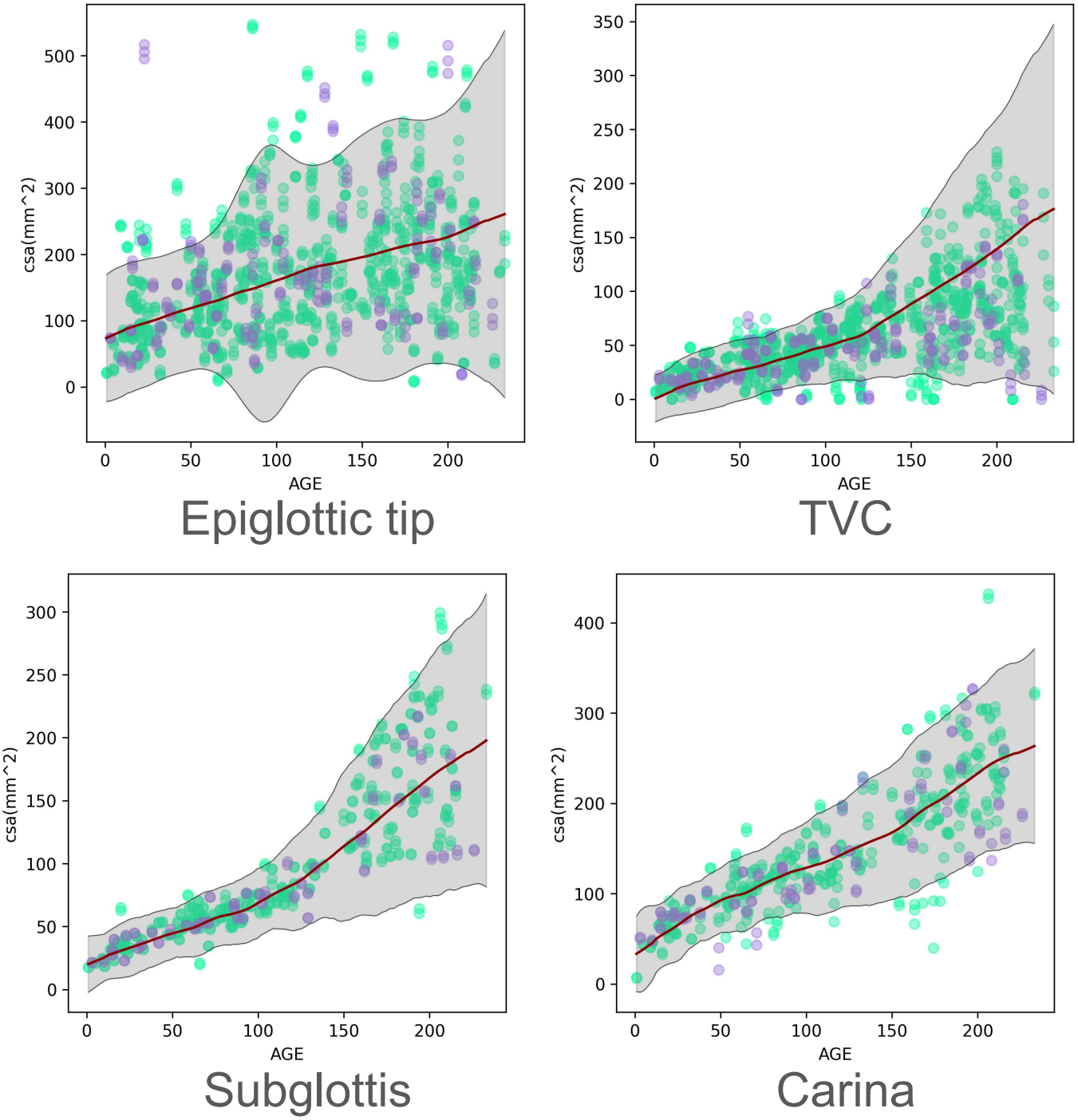

Figure 2: Visualization of learned $p(y|c_i, x)$ at differnet landmarks with `LucidAtlas`.