Master of Science
Thesis
VT2023

Optimization of b-value Acquisition and Parameter Estimation of Intravoxel Incoherent Motion Measurements

---

Ivan Rashid

Supervisors

Patrik Brynolfsson, Christian Jamtheim Gustafsson, and Lars E Olsson

Medical Radiation Physics, Lund
Faculty of Science
Lund University
www.msf.lu.se

Populärvetenskaplig Sammanfattning

Strålterapi är en vanligt förekommande behandling som ges till hälften av cancerpatienterna i landet. I fallet för prostatacancer, som utgör den vanligaste diagnosen, beror ofta cancerrelaterade dödsfall på återkomst av sjukdomen efter genomförd behandling. Vår hypotes är att detta kan minskas genom att identifiera strålrestistenta volymer i prostatan och öka stråldosen till dessa. Tyvärr finns det idag ingen pålitlig och icke-invasiv avbildningsteknik i klinisk drift för att kartlägga dessa områden. Enligt kända kunskaper inom strålningsbiologi beror strålinducerade cellskador på närvaro av syre. Därmed skulle brist på syre, det vill säga hypoxi, kunna agera som underlag för att identifiera dessa strålresistenta områden.

I en nyligen publicerad studie föreslogs en typ av magnetkameraundersökning för kartläggning av hypoxi. Bildtagningstekniken manipulerar och kodar in signalen från magnetkameran så att den blir känslig för slumpmässiga mikroskopiska rörelser i kroppen. Dessa rörelser kan exempelvis vara vattenmolekylernas slumpmässiga rörelse orsakad av rörelseenergi, eller blodflödet i slumpmässigt orienterade mikrokapillärer. För att utvinna den här informationen behöver mätningar göras med flera olika styrkor på inkodningen, då dessa rörelser är i olika storleksordningar. På så sätt kan dessa mätningar användas för att samtidigt avbilda både blodkärlstäthet och vävnadens mikrostruktur, såsom celltäthet. Det sistnämnda är möjligt då vattenmolekylernas rörelser begränsas av cellmiljön de rör sig i.

I studien som föreslår denna typ av undersökning för hypoxikartläggning användes informationen om blodflöde och celltäthet för att ta fram hypoxikartor enligt tillgång och efterfrågan. Tillgången utgörs av blodflödet, som består av syrebärande blodkroppar. Efterfrågan utgörs av celltätheten, alltså antalet celler som konsumerar syret. Med dålig cirkulation och hög celltäthet kan alltså hypoxiska områden identifieras, vilket ofta kan vara fallet för tumörer i prostatan.

Bildtagningstekniken som används för att ta fram underlaget för mikrocirkulationen och celltätheten är tidigare känd för dålig reproducerbarhet, till exempel på grund av signalbrus. För att lösa detta problem måste två frågor besvaras. Den första är vilka styrkor på inkodningen som är optimala att använda. Detta är relevant då olika vävnadstyper kräver olika inkodningsstyrkor för att karakterisera dem, vilket betyder att optimering behöver göras för varje organ för sig. Den andra frågan är vilken analysmetod som ska användas för att utvinna informationen om vävnaden från mätningarna. Brus är ett ständigt problem, och olika analysmetoder är olika känsliga för det. Det är vidare inte alltid uppenbart vilken metod som är bäst. Syftet med arbetet var att undersöka dessa frågor, med fokus på tillämpning i prostatan.

I arbetet tillämpades simuleringar och patientmätningar för att besvara dessa frågor. Simuleringarna användes för att utvärdera olika mätprotokoll och analysmetoder kvantitativt, med fokus på att studera osäkerheterna i informationen som erhålls om prostatans mikrocirkulation och celltäthet. Simuleringarna verifierades kvalitativt av patientmätningarna, där önskan var att kunna dra motsvarande slutsatser.

Resultaten av simuleringarna antydde att ett mätprotokoll vi optimerat kan prestera bättre än ett typiskt sådant som använts tidigare. Detta var inte lika tydligt i patientmätningarna, vilket skulle kunna bero på att de var för få. Mer arbete krävs för att dra säkrare slutsater. Gällande analysmetod kunde vi bekräfta och stärka med både simuleringar och patientdata att en nyligen föreslagen analysmetod presterar bättre än konventionella sådana som används idag. Vi har därmed visat att resultaten kan förbättras med befintlig data. I synnerhet de metoder studien som förelog denna typ av hypoxikartläggning använde. Detta tyder på att de föreslagna hypoxikartorna som redan verkar lovande skulle kunna förbättras genom att använda denna nya analys.

Abstract

Radiotherapy is commonly used for cancer treatment, given to approximately half of all cancer patients in Sweden. Despite its widespread usage, there is currently no non-invasive and reliable imaging method that can identify regions of radiation resistance. These regions may be identified by hypoxia, given that radiobiological research has demonstrated a significant correlation between oxygen levels and the extent of radiation-induced cellular damage.

A recent study has shown that magnetic resonance imaging (MRI) of intravoxel incoherent motions (IVIM) using diffusion-weighted MRI of the prostate shows promise for hypoxia mapping. The basis of these hypoxia maps are the IVIM parameters perfusion fraction and diffusion coefficient.

IVIM is recognized for being sensitive to noise and showing low reproducibility of parameter values. Existing literature highlights two issues that need to be resolved: optimization of image acquisition and reliable parameter estimation. The aim of this work is to address these issues by optimizing an acquisition protocol and to evaluate two new fitting algorithms; MIX and TopoPro; by comparing their estimation errors with commonly used algorithms to date.

Protocol optimization was performed with respect to diffusion weightings (b-values), and the number of averages as free parameters. The Cramér-Rao Lower Bound (CRLB) was used to provide estimates for the variances of the IVIM parameters for a set of b-values. Numerical minimization of the CRLB was employed to find optimal sets of b-values. Three protocols with different properties were produced and evaluated on a wide variety of simulated IVIM signals with respect to accuracy and precision. Comparison was made to a generic IVIM protocol typically found in literature. The best performing protocol was used in an in-vivo evaluation on the prostate and compared to the generic protocol with respect to estimation convergence. The optimized acquisition of in-vivo data improved convergence where the generic protocol otherwise would have failed to provide estimates due to noisy data. The optimized protocol increased the number of converged voxels by up to 20% and was found to vary between fitting algorithms. In conclusion, an optimized acquisition protocol may yield improved parameter estimates.

The evaluation of fitting algorithms was performed on simulated IVIM signals for three signal-to-noise ratios, where the signals were produced using a wide range of IVIM parameter combinations. MIX and TopoPro were compared to four conventional algorithms. Three of the conventional algorithms were found to be the most precise, but less accurate in certain parameter limits. MIX was found to be both precise and accurate. Furthermore, MIX showed a smaller dependence on the acquisition scheme than the others, both in simulations and in-vivo. MIX is therefore recommended as it is accurate, precise, and reduces the acquisition-dependency of IVIM parameter estimates. Furthermore, the usage of the Linear fit is not recommended as it showed the lowest accuracy, and lowest precision.

List of Abbreviations and Acronyms

MRI	Magnetic resonance imaging
DCE-MRI	Dynamic contrast-enhanced MRI
PET	Positron emission tomography
IVIM	Intravoxel incoherent motion
dMRI	Diffusion MRI
MIX	Microstructure imaging of crossing, fitting algorithm developed by Farooq et al.
TopoPro	Fitting algorithm that utilizing topological optimizers and variable projection, developed by Fadnavis et al.
b-value	Quantity describing the diffusion weighting
ADC	Apparent diffusion coefficient
D	Diffusion coefficient
D^*	Pseudo-diffusion coefficient of blood
f	Perfusion fraction of the signal
S_0	Signal from no diffusion weighting
NEX	Number of excitations
AI	Artificial intelligence
sIVIM	Simplified IVIM
CRLB	Cramér-Rao lower bound
SNR	Signal-to-noise ratio
Constant single tissue	Protocol with few b-values and high SNR. Optimized using the constant variance matrix on a single parameter set
Constant general	Protocol with few b-values and high SNR. Optimized using the constant variance matrix on ranges of parameter values
NEX general	Protocol with many b-values and lower SNR. Optimized using the NEX-scaled variance matrix on ranges of parameter values
RMSE	Root-mean-squared error

Contents

1 Introduction

Approximately half of all cancer patients in Sweden receive radiotherapy as part of their treatment [1], where prostate cancer constitutes the most common diagnosis [2]. The basic principle of radiotherapy is to cause lethal DNA damage to the tumour cells [3]. Research in radiotherapy and radiobiology has shown that a well oxygenated environment is an important factor for effective treatment [4]. Sparsely ionizing radiation; such as photons, which are often used in radiotherapy; rely on indirect damaging mechanisms that are dependent on radiation-oxygen interaction [5]. For poorly oxygenated environments (hypoxic), the delivered radiation dose may have to be up to three times larger in order to achieve the same effect as in well oxygenated environments [5]. Current clinical treatment planning protocols do not account for this effect.

This motivates the importance of images with hypoxia-dependent contrast in the planning stages of radiotherapy treatments. There are a number of potential techniques and modalities for both direct and indirect measurements of hypoxia, such as dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and positron emission tomography (PET) [6]. In the case of prostate cancer, DCE-MRI is not specific enough to be used as a biomarker for hypoxia. Furthermore, the hypoxia-tracers used in PET have failed to detect hypoxia in the prostate [6]. These methods are also invasive and require injection of contrast substances to work.

A non-invasive MRI technique that shows promise is intravoxel incoherent motion imaging (IVIM), a diffusion-MRI (dMRI) technique that can provide information on tissue perfusion without the use of contrast agents. In a study by Hompland et al., the creation of hypoxia maps using IVIM was conceptualized and validated [6]. They showed that the joint use of IVIM parameters that correlate with tissue perfusion and density could be linked to hypoxia scores derived from histology. They also established correlations between IVIM parameters and histological ones. All in all making IVIM imaging a candidate for hypoxia mapping.

The question remains on the reliability of these maps, as IVIM is associated with its own challenges and has historically shown ”modest” reproducibility [7]. Review articles on IVIM consistently point out two issues [7–13], the first being how to optimally acquire data for parameter estimation. IVIM signal curves differ from organ to organ depending on their tissue properties (Figure 1a). Therefore, acquisition protocols must be optimized for specific organs. The second issue is large uncertainties in the parameter estimations due to the analysis being sensitive to noise. In addition, there is no consensus on how to analyse the data (Figure 1b).

The purpose of this work to investigate these issues and attempt to improve on previously reported results. The first aim is to produce a clinically viable and optimized dMRI acquisition protocol for prostate imaging. The second aim is to evaluate two new analysis methods for IVIM, MIX and TopoPro; and compare them to more conventionally used methods, with regard to bias and noise sensitivity.

2 Theory

2.1 Diffusion-weighted Magnetic Resonance Imaging

2.1.1 Molecular Diffusion

Diffusion in MRI refers to the Brownian motion of water molecules, which are random translational motions on a microscopic scale caused by thermal energy [14]. This should not be confused with diffusion that refers to concentration equalization. The magnitude of the random motions can be described by a molecular diffusion coefficient D_m, which for water at 25 °C is 2.3 $\mu\mathrm{m}^2\,\mathrm{ms}^{-1}$ [14].

This diffusion coefficient is applicable only if self-diffusion of water is occurring, i.e. all of the displacement is caused by the random motion. External factors can affect the net displacement of bulk water. If the water is flowing, the diffusion coefficient will appear to be larger as the mean square displacement is increased due to translational motion caused by the flow [14]. Furthermore, if the water is restricted to

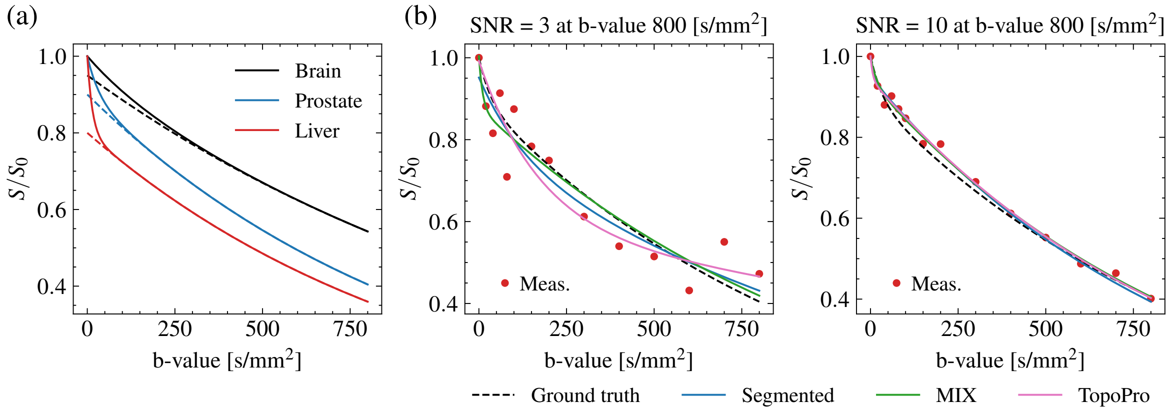


Figure 1: (a) Different types of tissue may yield different signal curves, as they depend on tissue perfusion and cell density. The liver is an example of an organ that is well perfused, while the brain is less so, and contains dense nerve networks that restrict the diffusion of water. (b) Different analysis methods may yield different results based on their accuracy and precision. Examples are shown for three of the six methods evaluated in this work.

spaces smaller than the mean squared displacement caused by the random motions, the diffusion coefficient will appear smaller, as the size of the motions are restricted to the bounds of the space [14]. Because of these effects, MRI-measured diffusion coefficients are called ”apparent diffusion coefficients” (ADC).

Therefore, by measuring molecular diffusion, one may be able to probe these external influences on the value of the diffusion coefficient. This is of great interest as these influences contain information on tissue properties. High ADC, i.e. measured diffusion coefficients that are larger than that of free water can be caused by tissue perfusion and other flowing liquids. Lower ADC can be caused by confined spaces in the tissue microstructure. Over the years, many techniques and applications have been developed, including but not limited to more advanced experimental techniques [15].

2.1.2 Diffusion Weighting

Diffusion-weighted MRI sensitizes the MRI signal to the molecular diffusion of water by encoding the signal in a way that introduces motion-dependent attenuation. This is conventionally achieved using the Stejskal-Tanner encoding technique [15]. Briefly, the encoding causes a signal loss which is reversed after a certain time interval. If water molecules have changed position during this time interval, the compensation will not be perfect, causing a motion-dependent signal attenuation [16].

The scale of motions that affect the magnitude of the signal depends on the strength of the diffusion encoding. This encoding strength, or diffusion weighting, is summarized by a quantity called the b-value, which depend on the duration, amplitude, and time spacing between the application of the encoding magnetic field gradients [14]. The b-value determines the attenuation caused by various degrees of incoherent motions. Signal contributions of water molecules with high apparent diffusion coefficients dissipate quickly, while signals from restricted water molecules remain for higher diffusion weightings.

2.2 Intravoxel Incoherent Motion

The imaging of intravoxel incoherent motions using dMRI was first presented in the 80s by Le Bihan, where it was noted that the ADC was found to be larger than expected in-vivo [14, 16]. This led to the theory that motions other than the random walk of water contributed to the de-phasing of the signal, such as blood microcirculation [16]. The theory is that the random architecture of the microvasculature causes a temporal incoherence in the flow during the application of the diffusion encoding gradients; causing the diffusion-like signal loss with a pseudo-diffusion coefficient D^* which is larger than the ADC of water molecules that are not affected by fluid motion.

The molecular diffusion coefficient D_m that describes the random motion of molecules can be estimated as

$$D_m = \frac{\bar{l}\bar{v}}{6}, \tag{1}$$

where $\bar{l}$ is the mean length of molecular jumps, and $\bar{v}$ the mean molecular velocity. For water, $\bar{l}$ is typically in the order of $0.1\ \mathrm{nm}$, and $\bar{v}$ in the order of $100\ \mathrm{m\,s^{-1}}$, resulting in typical values of D_m in the order of $1\ \mathrm{\mu m^2\,ms^{-1}}$. Le Bihan used eq. (1) to estimate the diffusion coefficient for microcirculation, using a model of randomly bent and oriented capillaries. In his estimation $\bar{l}$ was the mean capillary segment length between each bend ($60\ \mathrm{\mu m}$), and $\bar{v}$ the mean velocity of blood ($2\ \mathrm{mm\,s^{-1}}$). The "pseudo-diffusion coefficient" D^* was estimated to be in the order of $10\ \mathrm{\mu m^2\,ms^{-1}}$ [16]. Therefore, the randomly oriented blood flow would contribute to the de-phasing of the signal with a diffusion coefficient an order of magnitude larger than what is assumed for water, enabling the possibility of separating the coefficients. The following bi-exponential model was proposed for the dMRI-signal which includes a component for the perfusion effect, or in a stricter sense, all intravoxel incoherent motions:

$$S(b) = S_0 \left[f e^{-bD^*} + (1-f)e^{-bD} \right], \tag{2}$$

where S_0 is the unweighted signal, f the perfusion fraction of the signal, b the b-value that characterizes the strength of the diffusion encoding; which determines the signal attenuation caused by various degrees of incoherent motions. With this signal equation, the two diffusion coefficients can be separated if the bi-exponential behaviour is captured by the diffusion weighted acquisitions.

2.3 Interpretation of IVIM Parameter Maps

In a study aiming to produce hypoxia maps by Hompland et al., it was noted that a smörgåsbord of different methods and modalities have been tried for hypoxia imaging of the prostate, but have either failed or are challenging to perform [6]. Hompland et al. proposed a method to produce hypoxia maps based on maps of oxygen consumption and supply. In this case, the supply would be proportional microvasculature, and the consumption would be proportional to cell density. IVIM shows promise for this kind of imaging, as it conveniently provides parameters for such an analysis. It does however come with its own set of challenges (see Section 2.5).

Hompland et al. showed via histology that for the prostate, f can be correlated with blood vessel density, and D with cell density [6]. While correlation between D and cell density has been shown previously [7, 15], the correlation between f and microvessel density still needs to be established for various organs [7]. The correlation has currently been shown for the prostate, rectal cancer, meningioma, pancreatic carcinoma, and primitive neuro-ectodermal tumours [7].

The D^* parameter is a measure of the velocity of the blood flow in units of diffusion [17]. Therein lies the assumption that for the IVIM model of eq. (2) to be applicable; the movement of the blood needs to mimic a diffusive process during the diffusion encoding time, i.e. that the flow changes direction multiple times. This assumption is not necessarily true, hence Le Bihan proposed another IVIM model alongside the diffusive model of eq. (2); a model which assumes a ballistic blood flow, i.e. no directional changes during the diffusion encoding time. The diffusive and ballistic IVIM models can be considered as two extremes of temporal incoherence, where the diffusive represents full temporal incoherence, and the ballistic no temporal incoherence. Between these two extremes lies an intermediary regime for which no analytical expression of the signal equation currently exist [18].

Numerical simulations have shown that the flow has to change direction at least four times for full temporal incoherence to be achieved, making the diffusive model applicable [18–20]. This assumption has been shown to be violated in the brain [21, 22], liver [18, 19, 23], pancreas [18], and placenta [24, 25], where the flow has been either ballistic or intermediary (non-ballistic, but not temporally incoherent enough to be diffusive). These results warrant an investigation into the temporal incoherence of the perfusion in any organ for which IVIM measurements are of interest, as using the wrong model could lead to biases in the parameter estimates [19]. It is however outside the scope of this thesis, where only the diffusive IVIM model is considered.

2.4 IVIM Parameter Values in the Prostate

He et al. published a meta-analysis on IVIM imaging of the prostate, pooling the results from 20 studies and over 400 subjects [26]. The D parameter is suggested to range up to $1.7\ \mathrm{\mu m^2\,ms^{-1}}$, f up to 50%, and D^* varying greatly up to $50\ \mathrm{\mu m^2\,ms^{-1}}$, but is reportedly mostly under $10\ \mathrm{\mu m^2\,ms^{-1}}$ [26]. Forest plots show a difference in D between cancerous and non-cancerous tissue, where it is lower in cancerous tissue; thus indicating increased cell density in tumours [26]. The differences in D^* were also found to be significant; but as pointed out, suffers from large confidence intervals in its estimation. The differences in f were found to be insignificant, but unlike D^*, its estimation does not suffer from large uncertainties.

Furthermore, a multicentre study [27] performing daily IVIM imaging of patients undergoing radiotherapy on 1.5T MR-Linac systems reported pre-treatment values of f and D^* in general agreement with the meta-analysis by He et al. As the aforementioned study by Kooreman et al. points out, the reported D interval by the meta-analysis is large; however many studies places it in the range of 0.8-1.0 $\mathrm{\mu m^2\,ms^{-1}}$ for tumours, and 1.2-1.7 $\mathrm{\mu m^2\,ms^{-1}}$ for non-cancerous tissue [26]. Kooreman et al. did report a slightly larger D for tumours, but this could be attributed to the relatively low b-values used (maximum $500\ \mathrm{s\,mm^{-2}}$).

In conclusion, parameter values for the prostate are uncertain due to the large presented intervals. There are however plausible intervals that can be used for optimization and sanity checks.

2.5 The Challenges of IVIM

Reviews on IVIM consistently point out two issues that have to be resolved [7–13]. Namely, that there is no consensus on optimal acquisition protocols, and no consensus on data processing. This work aims to tackle both of these issues.

2.5.1 Optimization of Data Acquisition

Optimization of dMRI acquisition protocols for IVIM parameters are often discussed in relation to their repeatability and reproducibility [7]. Several reviews suggest that a larger number of b-values should be used (> 10) in order to accurately estimate the IVIM-parameters [11, 13]. These recommendations are partially based on the findings by ter Voert et al. [28], where the errors of the parameter estimates were found to be decreasing by adding additional b-values. However, one limitation of that study was that it did not account for the SNR of the diffusion-weighted images. Clinically, one is often restricted by the time it takes to perform the MRI measurements. Therefore, it may be more relevant to investigate different b-value schemes in relation to resource allocation, thus including the number of averages for each b-value in the optimization. A dMRI protocol with fewer b-values should therefore include more averages to make the comparison fair. It is therefore one of the aims of this work to investigate this aspect of b-value optimization.

Simultaneous optimization of b-values and their number of excitations (NEX) has been explored by others, and has produced improved parameter estimations compared to generic b-value schemes in the liver [29], brain [30], and kidney [31]. In a review by Federau, it is acknowledged that one also has to consider the SNR of the images, which is accounted for in this kind of optimization by allowing a variability in the number of averages for each b-value [8]. Furthermore, Federau mentions that minimization of motion and susceptibility artefacts is of importance, of which the former could be minimized by shortening the measurement time; warranting an optimized protocol based on measurement time, which can be partially achieved by reducing the number of b-values.

On the other hand, there seems to be more agreement regarding clustering of b-values versus uniform distributions, meaning that there are certain intervals on the b-scale that benefits from multiple tightly sampled measurements. Lemke et al. showed through simulations that when creating a very large set of b-values, where individual b-values were iteratively added to the set based on their contribution to lowering the estimation errors of the IVIM parameters, results in a frequency distribution of b-values where a large portion of the b-values are below $200\ \mathrm{s\,mm^{-2}}$ [32]. It is recognized that low b-values are important for accurate estimation of D^*, where many studies employing a larger number of b-values places a greater portion of them in the sub-200 $\mathrm{s\,mm^{-2}}$ range [11, 13]; implying that there is a general agreement on what could be regarded as the threshold where the IVIM effect becomes a negligible part of the signal [10]. This threshold is however organ-specific, which can be seen in Figure 1a, where the perfusion signal becomes negligible at different b-values for different organs. Wurnig et al. [33] showed that the optimal b-value threshold for well perfused organs could be as low as $20\ \mathrm{s\,mm^{-2}}$, where the threshold depended heavily on D^*.

The optimal b-value distributions are also organ specific, as concluded by Lemke et al., where they showed that well perfused organs require a different b-value distribution than low-perfused organs for optimized parameter estimation [32]. This complicates IVIM even further, as optimized protocols cannot simply be transferred between organs without thought and evaluation.

2.5.2 Parameter Estimation

Regarding the processing of the dMRI-images, most studies employ the segmented fitting strategies, which are described in Section 2.6.1 [7, 17]. This is due to the fitting problem being ill-conditioned and prone to failure [34]. For curve fitting to succeed, the data points have to be able to show bi-exponential behaviour, and the fitting algorithms need to be able to recognize this. This becomes increasingly difficult as f approaches 0% and D^* approaches D. Therefore, simply performing a non-linear least-squares fit to eq. (2) will result in noise-sensitive estimates due to the flexibility of the model and its parameters [10, 29].

Alternative strategies are continually being developed and evaluated; such as Bayesian estimation and AI-based strategies [7]. One of the aims of this work is to continue the evaluation of newly proposed methods and compare them to the conventional ones.

2.6 Fitting Strategies and Alternative IVIM Models

As previously mentioned, one cannot simply estimate the IVIM parameters by fitting eq. (2) to images acquired at multiple b-values. The flexibility of the model and its parameters render the results very sensitive to noise [10, 29]. Many alternative strategies have emerged over the years, where some overcome the previously mentioned issue of ambiguity by forcing the algorithm to consider two separate exponentials. In this work, six algorithms are evaluated. These are further categorized and explained below.

2.6.1 Segmented Bi-exponential Methods

Popular strategies used today are based on segmentation, meaning that parameters are fit to different ranges of b-values. In this work, two such methods are investigated. The first is what we call the *Subtracted* algorithm as described by Le Bihan [10]. The first step consist of fitting a mono-exponential defined as $S = S_0^{\mathrm{diff}} e^{-bD}$ to high b-values. Subsequently, a fit of $S = S_0^{\mathrm{perf}} e^{-bD^*}$ is performed to the residual signal after subtraction with the results of the first fit. Finally, f is estimated as $f = S_0^{\mathrm{perf}}/(S_0^{\mathrm{perf}} + S_0^{\mathrm{diff}})$. This method is graphically explained by Figure 2.

Another segmentation-based algorithm is the popular *Segmented* method, which in this work is implemented as described in the DIPY documentation [35]. The segmented method usually consists of fitting D to b-values above a specified threshold, followed by a fit to the full bi-exponential IVIM model keeping D fixed. The algorithm described by DIPY involves a few more steps. Before the fit to the bi-exponential, D^* is fit to b-values under a specified threshold. The acquired S_0 estimates are used to

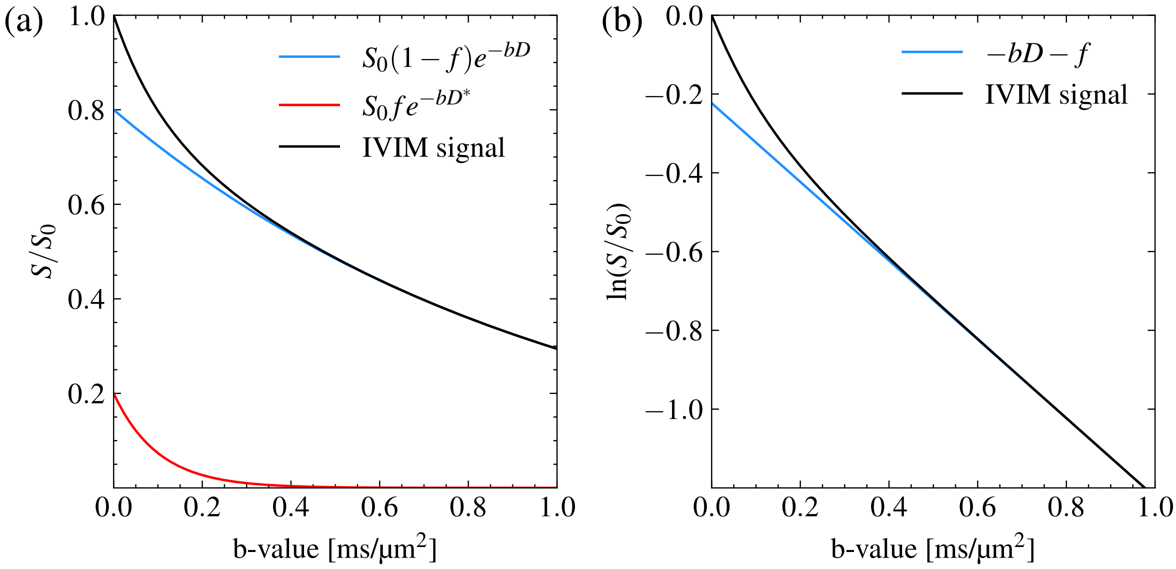


Figure 2: Graphical explanation of the fitting strategies. (a) The segmented strategies. The blue mono-exponential is fitted to high b-values. This is followed by fitting the perfusion-based mono-exponential, which serves as the second component to the IVIM signal. Note that $S_0^{\text{diff}} = S_0(1-f)$ and $S_0^{\text{perf}} = S_0 f$. (b) The linear fit shown in the logarithmic space, where the intercept with the y-axis approximates to f. In both (a) and (b), an f-value of 20% was used to simulate the signal.

estimate f as in the Subtracted method. Finally, keeping D fixed and using the calculated value of f as an initial guess, the full bi-exponential IVIM model is fit to the data.

2.6.2 Simplified Mono-exponential Models

The simplified models presented here are models that disregard the D^* parameter. This allows the IVIM-parameters f and D to be estimated using only three b-values [10]. It is indeed attractive from a clinical perspective as it reduces scan time, however another motivation for disregarding D^* is its poor reproducibility and sensitivity to noise [36]. Despite this, it shows promise in the diagnosis of prostate cancer [26]. However, for the purpose of producing hypoxia maps as done in the Hompland study [6], D^* is not needed, making these methods viable.

The first simplified method is the so-called *Linear fit*, where a line is fitted to $\ln(S/S_0)$ for b-values over a certain threshold [6]. This threshold is placed where it is assumed that no perfusion signal remains (generally between 200 and 600 $\mathrm{s\,mm^{-2}}$) [10]. The linear fit gives an estimation of D. The estimate of f is obtained by extrapolating the line to the intercept with the y-axis, as the intercept is approximately $-f$ [10].

The second simplified model evaluated in this work is the creatively named *simplified IVIM*-model (sIVIM). It is an exponential model defined as

$$S(b) = S_0\left[\delta(b)f + (1-f)e^{-bD}\right], \tag{3}$$

where δ is the discrete delta yielding 0 for all $b \neq 0\,\mathrm{s\,mm^{-2}}$, and $= 1$ for $b = 0\,\mathrm{s\,mm^{-2}}$ [29]. The difference between the two simplified models is that sIVIM includes $b = 0\,\mathrm{s\,mm^{-2}}$ in a non-linear least squares fit instead of relying on an extrapolation to estimate f. Similar to the Linear fit, with the exception for b-value $0\,\mathrm{s\,mm^{-2}}$, the b-values used should be above a threshold where it is assumed that contribution of perfusion is negligible.

2.6.3 Variable Projection

Variable projection is a problem formulation that has been proven to be a computationally efficient optimization method applicable to linear combinations of non-linear functions [37]. The principle is to separate the linear and non-linear parameters, reformulating the problem into a pure non-linear least-squares problem, the solution of which can be used to efficiently solve for the linear parameters [37]. One can note that the bi-exponential IVIM model (eq. (2)) is indeed a linear combination of the non-linear base functions e^{-bD^*} and e^{-bD}, hence making it possible to use this method for IVIM parameter estimation [34].

As described by Golub and Pereyra [37], in general, the residual of a linear combination of non-linear functions is given by

$$r_i(\vec{f}, \vec{\beta}) = y_i - \sum_{j=1}^{n} f_j \phi_j(\vec{\beta}; b_i), \tag{4}$$

where $\vec{y}$ is the vector of observations dependent on the independent variable values in $\vec{b}$; $\vec{f}$ the linear parameters; $\vec{\beta}$ the non-linear parameters. What we can immediately see here by the chosen notation is that in the case of IVIM, the perfusion and diffusion fractions are represented by $\vec{f}$, the b-values of the observations by $\vec{b}$. And the non-linear parameters in $\vec{\beta}$ would be D and D^*.

Translating eq. (4) into matrix notation, one gets

$$||r(\vec{f}, \vec{\beta})||_2^2 = ||\vec{y} - \Phi(\vec{\beta})\vec{f}||_2^2, \tag{5}$$

where Φ is a matrix with columns being the base functions evaluated for $\vec{\beta}$, and rows being the base functions evaluated for $\vec{b}$. This makes it a matrix whose number of columns correspond to the number of non-linear functions in the linear combination of functions, and number of rows corresponding to the number of observations.

The linear parameter vector can be determined by solving the linear least-squares problem

$$\vec{f} = \Phi(\vec{\beta})^{\dagger}\vec{y}, \tag{6}$$

where † denotes the Moore-Penrose inverse, defined as $\Phi^{\dagger} = (\Phi^T\Phi)^{-1}\Phi^T$ [38]. Substituting eq. (6) into eq. (5), the residual becomes

$$||r(\vec{\beta})||_2^2 = ||\vec{y} - \Phi(\vec{\beta})\Phi(\vec{\beta})^{\dagger}\vec{y}||_2^2, \tag{7}$$

which is independent of the linear parameters. Estimations for all parameters can be obtained by solving the non-linear least-squares problem in eq. (7) for $\vec{\beta}$, which can in turn be used to solve the linear least-squares problem in eq. (6) for $\vec{f}$ [37].

This problem formulation has been used in the proposal of two fitting algorithms for IVIM; MIX by Farooq et al. [38, 39]; TopoPro by Fadnavis et al. [34]. The difference between the two are the optimizers that are used to solve the minimization problems above. MIX employs a 3-step fit, where in the first step, a stochastic search using differential evolution is used to find $\vec{\beta}$ [39]; second step, making use of a convex optimizer for a constrained search for $\vec{f}$ with $\sum_{j=1}^{n} f_j = 1$; finalizing the process with a constrained trust-region solver on eq. (4) using the estimates from step 1 and 2 as initial guesses [38].

TopoPro also employs a 3-step fit, where a simplicial homology global optimizer (SHGO) is used to solve eq. (7), followed by a convex optimizer to solve eq. (6), finalizing with another SHGO-optimization of the full functional in eq. (4) using the previous results to tighten the constraints [34].

2.7 Cramér-Rao Lower Bound for b-value Optimization

Cramér-Rao optimization of acquisition protocols has shown to improve parametric maps for the liver [29, 40] and the brain [30].

The task of a function-fitting algorithm is to estimate a set of parameters such that the function, as precisely as the uncertainty of the observations allows it to, can describe the variation of the observations.

If one has knowledge of the probability density function of the stochastic noise, one can determine the Cramér-Rao lower bound (CRLB), i.e. the lower bound of the variances of the parameter estimates [41].

In the case of MRI, it is known that the noise is Rice distributed [42]. Therefore, there is an a priori knowledge of the probability density function of the noise. The CRLB of Rice-distributed noise has been derived, but can be simplified with the assumption that the Rice distribution becomes Gaussian at signal-to-noise ratios > 6 [41] (the approximation may even be valid at SNR > 3 [42]). This leads to the following formula for the Fisher information matrix, where the diagonal of its inverse is the lower bound for the variance of each parameter:

$$F_{i,j} = \sum_{n}^{N} \frac{1}{\sigma_n^2} \frac{\partial A_n}{\partial \beta_j} \frac{\partial A_n}{\partial \beta_i}, \tag{8}$$

where A_n is the function whose parameters are to be estimated for data point n, $\beta_{i/j}$ a parameter from the vector of parameters $\vec{\beta}$, σ_n^2 the variance of the signal for data point n [41]. In matrix notation, eq. (8) can be written as

$$F_{i,j} = \left(\frac{\partial \vec{A}(\vec{\beta})}{\partial \beta_i} \right)^T V^{-1} \left(\frac{\partial \vec{A}(\vec{\beta})}{\partial \beta_j} \right), \tag{9}$$

where V^{-1} is an inverse variance matrix, a diagonal matrix with the elements $1/\sigma_n^2$.

In the case of the bi-exponential IVIM model, A is eq. (2), which in eq. (9) is viewed as a function of all its parameters, i.e. $S(\vec{b}, S_0, f, D^*, D) = S(\vec{b}, \vec{\beta}) = \vec{S}(\vec{\beta})$. Note that the evaluation of $S(b, \vec{\beta})$ for each b-value yields $\vec{S}$; a vector whose length corresponds to the number of b-values N. Finally, by evaluating the partial derivatives of S, one can construct the Fisher information matrix, whose diagonal gives the lower bound of the variances for S_0, f, D^*, and D.

Now that the CRLB of each parameter can be determined for a set of b-values, one can define a cost function to minimize. Hence, making it possible to iteratively find a set of b-values that yields the lowest CRLB; therein using the Cramér-Rao lower bound for b-value optimization.

3 Materials and Methods

Protocols and algorithms were evaluated quantitatively using simulations, verified qualitatively by in-vivo data. The simulations and data analysis were performed using Python scripts (v. 3.9.12) implemented in Hero v. 2023.1.0 (Hero Imaging AB, Umeå, Sweden). Python scripts utilized NumPy v. 1.21.5, SciPy v. 1.9.3, and DIPY v. 1.5.0. Phantom and in-vivo images were acquired using GE SIGNA$^{\text{TM}}$ Architect 3T v. MR29, denoised using AIR$^{\text{TM}}$ Recon DL reconstructions (General Electric, Milwaukee WI) to improve SNR.

3.1 b-value Optimization Using the Cramér-Rao Lower Bound

To use the Cramér-Rao framework for variance estimation as explained in Section 2.7, information needed to be acquired regarding the inverse variance matrix, and whether the SNR of the acquired images would be sufficient to use the Gaussian noise approximation which eq. (9) is based on. This motivated the need for SNR measurements of our intended protocol parameters.

3.1.1 Measurement and Quantification of Signal- to Noise-Ratio and its NEX-dependence

Measurements were performed on a spherical water phantom to gather information for the inverse variance matrix in eq. (9). dMRI was performed for b-values 0, 50, 240, 800 $\mathrm{s\,mm^{-2}}$ with NEX 1, 3, 6, 9 for each b-value. Further measurements details are found in Section 3.4. SNR was quantified by applying a Gaussian filter to the image, and subtracting it from the original, which corresponds to subtracting low frequencies in the frequency domain [43]; resulting in a noise image (Figure 3). A square ROI placed in the centre of the sphere was used to extract the mean and standard deviation of the noise image, where

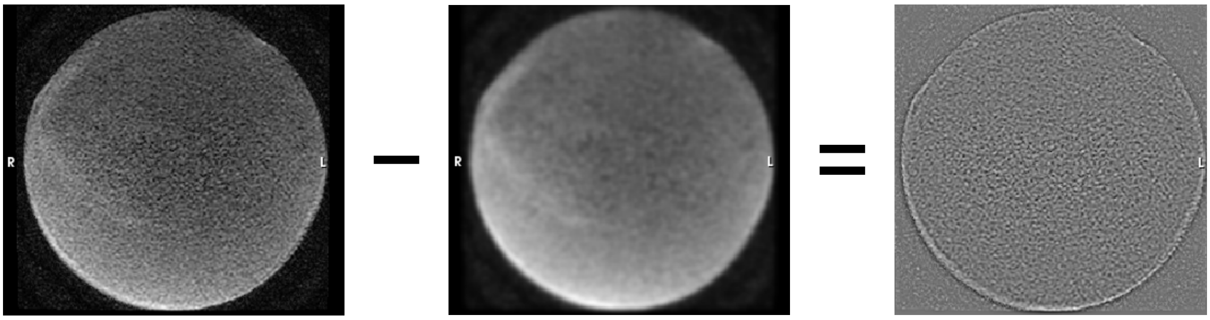


Figure 3: The quantification of SNR. A Gauss-blurred image was subtracted from the original, resulting in an estimated image of the noise. The SNR was determined by extracting the mean and standard deviation of the noise image. This method of SNR-quantification is described by McCann et al. [43].

image noise $=$ standard deviation$/\sqrt{2}$, and SNR $=$ mean/image noise [44]. The image noise was plotted as a function of NEX, shown in Figure 4. The noise is proportional to $1/\sqrt{\mathrm{NEX}}$. Furthermore, we validate the Gaussian approximation used in the Cramér-Rao theory (Section 2.7) by showing that SNR is at least 3 for a single excitation without noise reduction.

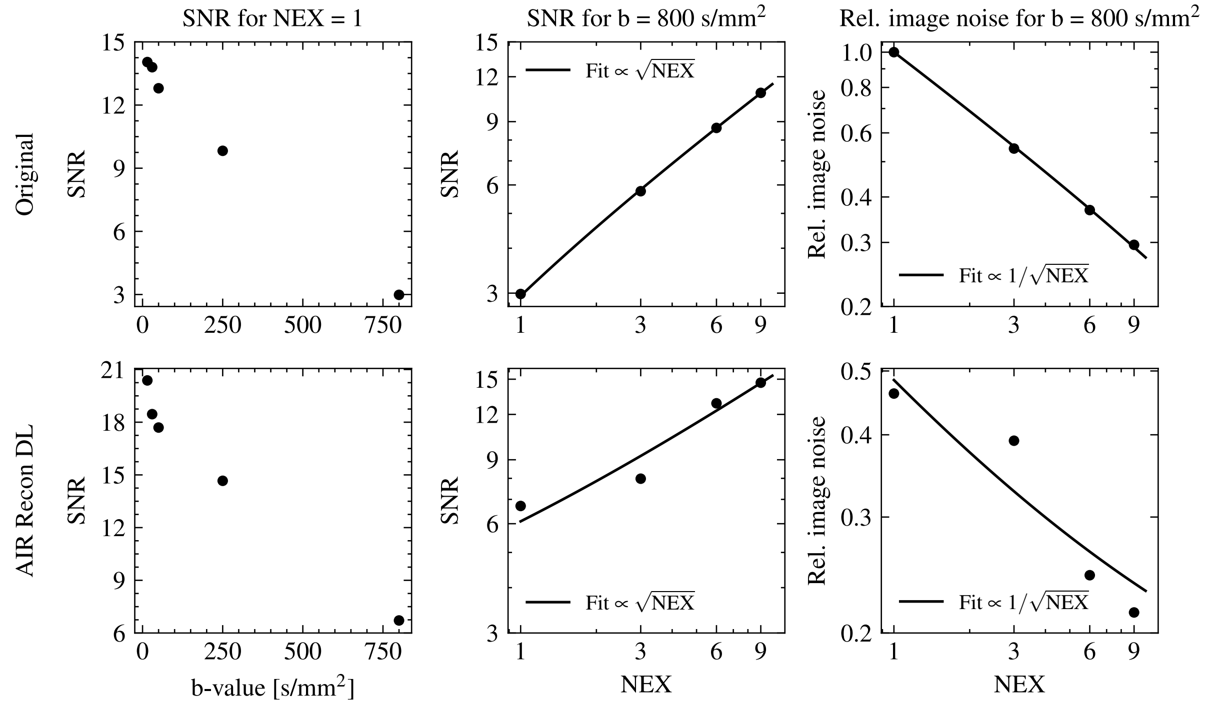


Figure 4: The results of the SNR vs. NEX measurements for both original and denoised data. The denoised images were reconstructed using AIR Recon DL. A clear NEX-dependency is observed, in-line with basic SNR theory.

3.1.2 The Noise Variance Matrix

Two different noise variance matrices were used in the performed optimizations. The purpose of using both of these matrices was to produce optimized protocols with different properties. These would be used in the investigation of whether the accuracy and precision of IVIM parameter estimation was better when using more diverse b-sets, or fewer unique b-values with higher SNR.

The first matrix was a constant (unit) matrix, assuming no difference in image noise across b-values.

Optimizations performed using this matrix are labelled as *Constant*. This was used to obtain optimized b-sets with as few b-values as possible.

The second variance matrix was one derived from the SNR-measurements above. Original image reconstructions showed a clear NEX-dependency in the SNR. The same dependency was not as clear for the AIR Recon DL reconstructions, which can be expected due to uncertainties in the noise estimation.

Since the noise was found to be proportional to $1/\sqrt{\mathrm{NEX}}$, the inverse variance matrix becomes a diagonal matrix of the NEX for that particular b-value. E.g. if the b-set consisted of two acquisitions of b-value $800\ \mathrm{s\,mm^{-2}}$, two of the diagonal elements would attain the value 2 in indices corresponding to those b-values in the b-value vector mentioned in Section 2.7. This variance matrix was used to obtain more diverse b-value sets (labelled as *NEX*), as a result of accounting for the change in SNR with increasing NEX.

3.1.3 Calculation of Fisher Matrix Elements

The partial derivatives of eq. (2) were derived to be

$$\begin{aligned}
\frac{\partial S}{\partial S_0} &= f e^{-bD^*} + (1-f)e^{-bD} \\
\frac{\partial S}{\partial f} &= S_0(e^{-bD^*} - e^{-bD}) \\
\frac{\partial S}{\partial D^*} &= -S_0 f b e^{-bD^*} \\
\frac{\partial S}{\partial D} &= -S_0(1-f) b e^{-bD}.
\end{aligned}$$

A matrix was constructed where each column was represented by a partial derivative, and each row the evaluation of the partial derivatives for a specific b-value. The matrix was used in eq. (9) with one of the variance matrices described above. The variance for each parameter was extracted from the diagonal of the resulting matrix, corresponding to the index of the partial derivative for that parameter.

3.1.4 The Objective Function

Similar to the CRLB-based optimizations done for the liver [29, 40], a cost function was defined as the sum of the coefficients of variation for each parameter, i.e.

$$\epsilon = \frac{\sigma_f}{f} + \frac{\sigma_{D^*}}{D^*} + \frac{\sigma_D}{D}, \tag{10}$$

where each standard deviation σ was acquired by determining the square root of eq. (9).

Minimization of eq. (10) was performed with `scipy.optimize.minimize` using the Nelder-Mead algorithm.

3.1.5 Optimization for Intervals of Parameter Values

The objective function described above only allows optimization for a specific set of parameter values. However, tissue would realistically consist of a range of parameter values. To account for this, an attempt to find a set of b-values optimized for several sets of parameters was made.

This was done by defining ranges for f and D^* using information from the meta-analysis discussed in Section 2.3. An array was constructed from the two ranges with 10 elements each, giving an array of 100 unique sets of f and D^*. D was fixed to $1.0\ \mathrm{\mu m^2\,ms^{-1}}$. Optimizations were performed for each array element, resulting in an optimized b-set for each element. For each acquired b-set, costs were calculated for every value of D^*, yielding cost maps dependent on b-set, D^*-value, and f-value. Summation of the costs for all values of D^* was performed, and the most optimal b-set was determined to be the one with the lowest total cost. Hence, an optimal b-set was obtained for each value of f.

Table 1: Summary of the objectives and variables used to produce the three optimized protocols.

Name	Variance matrix	Parameter set	Objective
Constant single tissue	Constant	Single parameter set specific to the prostate from He et al. [26].	Few unique b-values with more averages optimized for a single type of tissue.
Constant general	Constant	Ranges of parameter values.	Few unique b-values with more averages optimized for general acquisition.
NEX general	NEX-scaled	Ranges of parameter values.	Many unique b-values with fewer averages optimized for general acquisition.

The final step would be for the user to select the b-set of a desired value of f. In this work, optimal b-value schemes were finalized by observing a general pattern in the b-values for the different values of f, selecting a b-value scheme which represented a median in the occurrences of certain b-values.

3.1.6 Performed Optimizations

Three optimizations were performed. For all of these, an upper b-value limit was set to 800 $\mathrm{s\,mm^{-2}}$ to avoid signal dependecies at high b-values that are not accounted for in the bi-exponential IVIM model. Furthermore, the total number of data points was limited to 14, which corresponds to 14 NEX. The objectives of the optimizations are summarized in Table 1.

Firstly, a b-set was optimized for a single set of parameters using the constant variance matrix, in order to obtain a set of few unique b-values optimized for a single type of tissue. The parameters were set to $f = 0.1, D^* = 20\ \mathrm{\mu m^2\,ms^{-1}}, D = 0.5\ \mathrm{\mu m^2\,ms^{-1}}$. The purpose of the low D was to optimize the b-values for tumours. This set is denoted as *Constant single tissue*.

Secondly, an attempt was made to acquire a b-set with few unique b-values that was optimized for multiple types of tissues by using the constant variance matrix on ranges of parameter values, as described in Section 3.1.5. The ranges were set to $f \in [0, 0.5]$, $D^* \in [5, 30]\ \mathrm{\mu m^2\,ms^{-1}}$, with fixed $D = 1.0\ \mathrm{\mu m^2\,ms^{-1}}$. The acquired b-set is denoted as *Constant general*, where general refers to the set being optimized for multiple types of tissue.

Thirdly, the second optimization was repeated using the NEX-scaled variance matrix in order to acquire a diverse b-set optimized for multiple types of tissue, hence labelled as *NEX general*.

3.2 Quantitative Evaluation of Fitting Algorithms

3.2.1 Implementation of Fitting Algorithms

The fitting algorithms were implemented in Python using the NumPy, SciPy, and DIPY libraries. The code for all implementations can be accessed through GitHub (`https://github.com/IvanARashid/MSFT02-Masters-project-IVIM-optimization/`).

The Linear fit was implemented as described by Hompland et al. [6]. A linear function was defined and fitted to b-values greater than 200 $\mathrm{s\,mm^{-2}}$ in the logarithmic space, where the signals were normalized against the signal for b-value 0 $\mathrm{s\,mm^{-2}}$. The fit was performed by the scipy `optimize.lsq_linear`

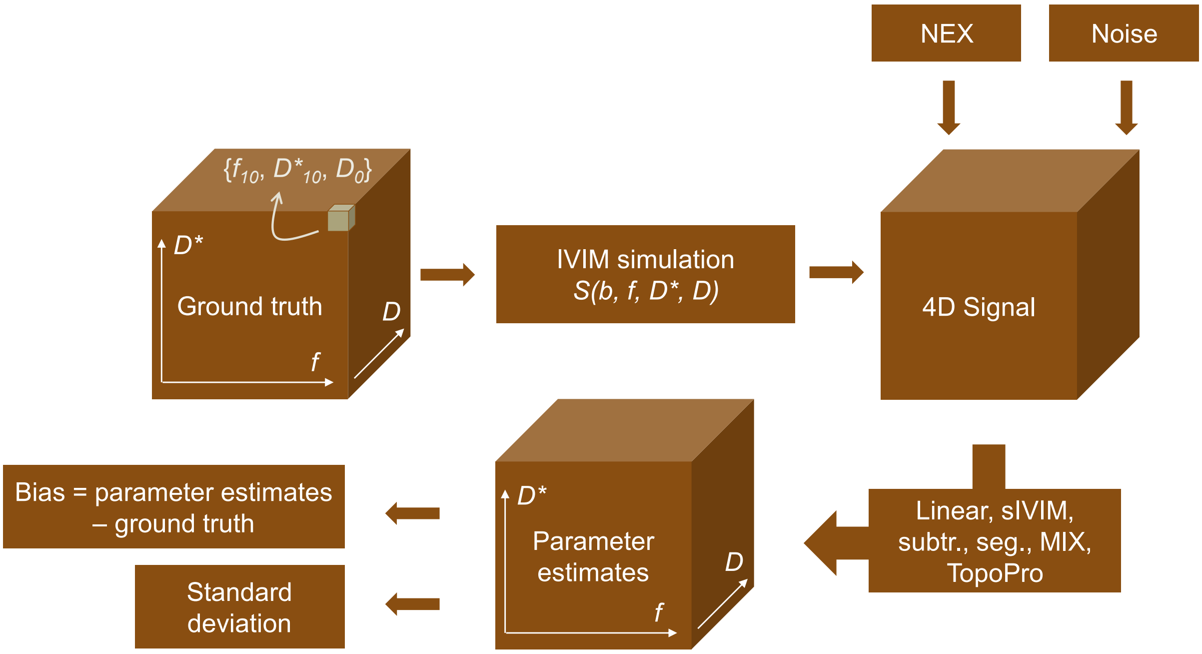


Figure 5: Schematic figure of the simulations. An array of parameter combinations was defined, where each element constituted a set of f, D^*, and D values. Noised signals were simulated for these, accounting for the number of averages of each b-value. This was followed by parameter estimation using the six algorithms.

function.

The sIVIM fit was implemented by using the scipy `curve_fit` to fit data to eq. (3). The b-values used were 0 and $\geq 200\,\mathrm{s\,mm^{-2}}$.

The Subtracted and Segmented fits were both implemented using the scipy `curve_fit` function for the steps described in Section 2.6.1. The upper b-value threshold for the perfusion fit was set to $100\,\mathrm{s\,mm^{-2}}$, and the lower b-value threshold of the diffusion fit $200\,\mathrm{s\,mm^{-2}}$.

For the variable projection methods, the DIPY-implementation of MIX was copied and slightly modified (`https://github.com/dipy/dipy/blob/master/dipy/reconst/ivim.py`). The modification was to loosen hard-coded bounds of the fits, as they were found to be too stringent. Similarly, TopoPro was taken from the authors GitHub (`https://github.com/ShreyasFadnavis/topopro/`), where the hard-coded bounds for the first level of SHGO were loosened. Furthermore, the number of SHGO iterations were increased from 2 to 5, as it provided more stable estimations. At 2 iterations, it was found to frequently fail and produce near-constant signal fits. This modification came with the downside of greatly increased computation time.

Bounds were set for all algorithms such that estimates were restricted to $f \in [0,1]$, $D^* \in [5,100]\ \mathrm{\mu m^2\,ms^{-1}}$, and $D \in [0,4]\ \mathrm{\mu m^2\,ms^{-1}}$.

3.2.2 Simulations

To account for various tissue properties, a total of 1000 parameter combinations ordered in a 10 × 10 × 10 array was defined; with each dimension corresponding to increasing f, D^*, and D, such that each element consisted of a unique parameter combination (Figure 5). Linear ranges were used, where $f \in [0.02, 0.5]$, $D^* \in [5,30]\ \mathrm{\mu m^2\,ms^{-1}}$, and $D \in [0,3]\ \mathrm{\mu m^2\,ms^{-1}}$. Signals were simulated using eq. (2) for the b-value scheme the simulations were performed for.

To investigate noise sensitivity, Rician noise was simulated 100 times for each parameter combination such that a specific SNR was obtained for $S(800\,\mathrm{s\,mm^{-2}})$, giving a total of 100 000 signal curves. The noise was applied by sampling real and imaginary components from a Gaussian distribution with standard deviation $\frac{S(800\,\mathrm{s\,mm^{-2}})}{\mathrm{SNR}}$, thus weighting the SNR for the highest b-value. The sampled real and imaginary noise components n_r and n_i were added to the signal S, followed by taking the magnitude, according to [45]

$$M = \sqrt{(S + n_r)^2 + n_i^2}. \tag{11}$$

For instances where multiple NEX of a certain b-value was simulated, noised signals corresponding to the number of NEX was generated and averaged in the complex domain before taking the magnitude.

The parameters were estimated by performing fits with the six algorithms to the 100×1000 signal curves. The 100 noised signals of each parameter combination were used to obtain a standard deviation of the estimates. The bias of the algorithms were determined by subtracting the ground truth from the estimates obtained from fits to the un-noised signal. The root-mean-square error (RMSE $= \sqrt{\mathrm{bias}^2 + \mathrm{standard\ deviation}^2}$) was used as a metric for total error. To study the RMSE trends as a function of an individual parameter, the RMSE of the planes made up of the two other parameters were summed along the aforementioned parameter axis.

3.3 Error Dependency on b-value Threshold and SNR for Segmentation-based Methods

To further understand the impact of the b-value threshold of the Linear, sIVIM, Subtracted, and Segmented methods on the RMSE, simulations were performed using the *Generic* protocol with b-values 250 and 350 $\mathrm{s\,mm^{-2}}$ added. The simulations were performed for SNR $\in [3, 21]$ with step size 3, and b-value thresholds $\in [100, 400]\ \mathrm{s\,mm^{-2}}$ with step size 50 $\mathrm{s\,mm^{-2}}$. The RMSE for all parameter combinations were summed for each parameter, and plotted against SNR and b-value threshold to investigate how SNR could affect the choice of b-value threshold when performing a fit using the conventional methods.

3.4 In-vivo Measurements and Evaluation

Diffusion-weighted images were acquired using a standard 3-directional SE-EPI sequence with FOCUS and linear shimming applied over the field-of-view. Twenty non-spaced slices with a thickness of 3 mm were acquired with matrix size 160×80 and in-plane voxel size of $1.5 \times 1.5\mathrm{mm}^2$, resulting in a field-of-view of 24×12 cm^2 in the frequency and phase encoding directions respectively. TE 62 ms, TR 4727 ms, with scan time approximately 3 minutes and 40 seconds for each protocol.

Multiple excitations of b-values were geometrically averaged in the image reconstruction by the scanner. For every protocol, the number of excitations for b-value $0\ \mathrm{s\,mm^{-2}}$ were multiplied by 3 to account for the fact that the non-zero b-values effectively get three excitations in different diffusion directions per input NEX on the GE MRI system, if three directional diffusion is enabled. The acquired images were reconstructed with AIR Recon DL for image noise reduction. Screenshots of the pulse sequence settings are presented in Appendix A.

A total of six patients were scanned with both the *Constant single tissue* and *Generic* b-value schemes. One patient was excluded from the evaluation due to severe susceptibility artefacts in the diffusion-weighted images. Whole-prostate and tumour regions of interest (ROI) were drawn by oncologists. Results of T2- and diffusion-weighted images from prior diagnostic examinations was taken into account for tumour contouring on non-diagnostic T2-weighted images acquired for radiation treatment planning.

The diffusion-weighted images were registered to the non-weighted image, which in turn was registered to the T2-weighted image which was the basis for the tumour ROI. These registrations accounted for inter-b-value motion. Diffusion-weighted images with motion artefacts were discarded from the curve fitting. Only one diffusion-weighted image was discarded for this reason.

Parameter estimates were obtained using the six algorithms with the same bounds and settings described in Section 3.2.1.

3.4.1 Protocol Evaluation

The MRI acquisition protocols were evaluated by quantifying their success in the depiction of bi-exponential signal behaviour. This was done by defining a percentage of success as the ratio of the number of voxels with parameter estimates within certain bounds divided by the total number of voxels in a whole-prostate ROI, i.e. $\text{Success} = \frac{\text{\# voxels depicting bi-exponential}}{\text{Total \# voxels}}$. The bounds were set by studying histograms of parameter estimates. Bounds were set such that voxels with $f \in [0.01, 0.9]$ and $D^* \in [5, 90]\ \mu\mathrm{m}^2\,\mathrm{ms}^{-1}$ were counted as successful bi-exponential fits. The success-metric was determined for f and D^* independently for both protocols. The quotient of these, i.e. $\frac{\text{Success opt.}}{\text{Success gen.}}$ was used to evaluate how well the optimized protocol performed compared to the generic. A value > 1 would indicate that the optimized protocol provided data that depicted the bi-exponential behaviour better than the generic protocol, while values < 1 would indicate that the generic protocol performed better.

3.4.2 Algorithm Evaluation

The algorithms were qualitatively evaluated by studying the probability densities of their voxelwise parameter estimates. Histograms of all patients were averaged for two regions: 1) of parameter estimates from voxels in an ROI surrounding the tumour and 2) from voxels in the prostate minus the tumour ROI. Both the tumour ROI and prostate ROI were drawn by oncologists.

4 Results

4.1 b-value Optimization Using the Cramér-Rao Lower Bound

The three optimized protocols described in Section 3.1.6 and a generic protocol are presented in Table 2. An example of the cost maps used to determine the optimal b-value scheme for a range of parameters is shown in Figure 6, where the b-set with the lowest sum along the D^*-axis was determined to be the most optimal for this particular combination of f and D. It can be noted that the value of the objective function is often large when D^* approaches the range of D.

As expected, the optimizations performed using the constant variance matrix yielded fewer unique b-values with a larger NEX allocated per b-value. This in comparison to *NEX general* and *Generic* which employ larger number of unique b-values with fewer NEX, and are therefore expected to yield lower SNR in the diffusion-weighted images. An overview and ranking of protocol performance is presented in Table 3. These ranks are based on the results of the following sections.

Table 2: The protocols used for evaluation of the Cramér-Rao optimization, and the evaluation of the fitting algorithms. Note that the optimized protocols originally allocated 1 NEX for 0 $\mathrm{s\,mm^{-2}}$. An extra excitation was added to obtain a preciser measurement of S_0. The sum of all NEX for each protocol is 15, yielding comparable measurement times.

Protocol	b-value$_{\mathrm{NEX}}$ $[\mathrm{s\,mm^{-2}}]$
Constant single tissue	$0_2, 50_3, 240_5, 800_5$
Constant general	$0_2, 80_3, 350_6, 800_4$
NEX general	$0_2, 60_1, 70_1, 220_1, 310_3, 350_1, 520_3, 800_3$
Generic	$0_2, 20_1, 40_1, 60_1, 80_1, 100_1, 150_1, 200_1, 300_1, 400_1, 500_1, 600_1, 700_1, 800_1$

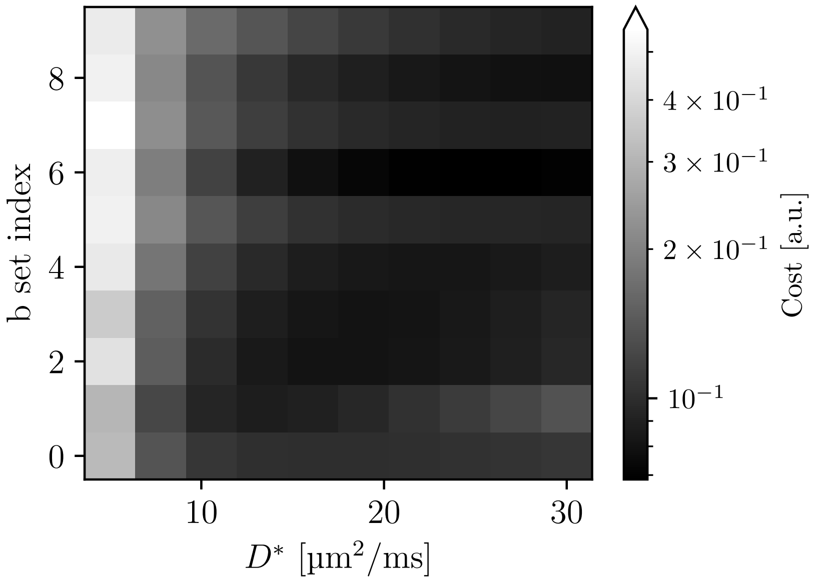


Figure 6: The value of the objective function used for protocol optimization (eq. (10)) for different b-value schemes and values of D^*, with fixed $D = 1.0$ µm^2 ms^{-1} and $f = 5\%$. The protocol with the lowest total cost was the one labelled as index 3, which in this report is called *NEX general*.

Table 3: Overview of qualitative ranking of protocol performance. The basis for these scores can be found in the Figures in Appendices B, C, and D. The protocols are ranked 1-4, with lower rankings being better.

Protocol	f bias	f RMSE	D^* bias	D^* RMSE	D bias	D RMSE	Sum
Constant single tissue	2	1	4	2	1	1	11
Constant general	1	2	1	4	2	2	12
NEX general	3	3	3	3	3	3	18
Generic	4	4	2	1	4	4	19

4.2 Quantitative Evaluation of Fitting Algorithms and Acquisition Protocols

Figure 7 shows the RMSE trends for f, D^*, and D respectively, at the lowest tested SNR of 3. Results are shown for the *Constant single tissue* and *Generic* protocols as they were the best and worst performing respectively. Results for all protocols and SNR are found in Appendix B.

For both f and D, the *Constant single tissue* protocol in combination with the conventional fitting strategies (barring the Linear fit) showed generally lower RMSE than the MIX and TopoPro combinations, along with any combination made with the *Generic* protocol. Although MIX showed a lower bias in f when D^* was low. For estimation of D^*, MIX in combination with the *Generic* protocol showed the best performance, although the *Constant single tissue* together with the sIVIM, Subtracted, and Segmented methods were not far off.

The RMSE trends give a quick overview of the best performing protocol-algorithm combinations, however to further understand the errors in the parameter estimates, the RMSE and bias maps presented in Figures 8-13 need to be studied. Results are only shown for $D = 1.0$ µm^2 ms^{-1} as the patterns were similar for all simulated D. Results for all ground truth values of D are presented in Appendices C and D.

Figure 8 shows that the RMSE of f approached the bias at SNR 10 for the conventional methods, while MIX and TopoPro required a larger SNR to approach their respective biases. It can be noted that sIVIM, Subtracted, and Segmented performed well at low SNR with an RMSE of approximately 15-

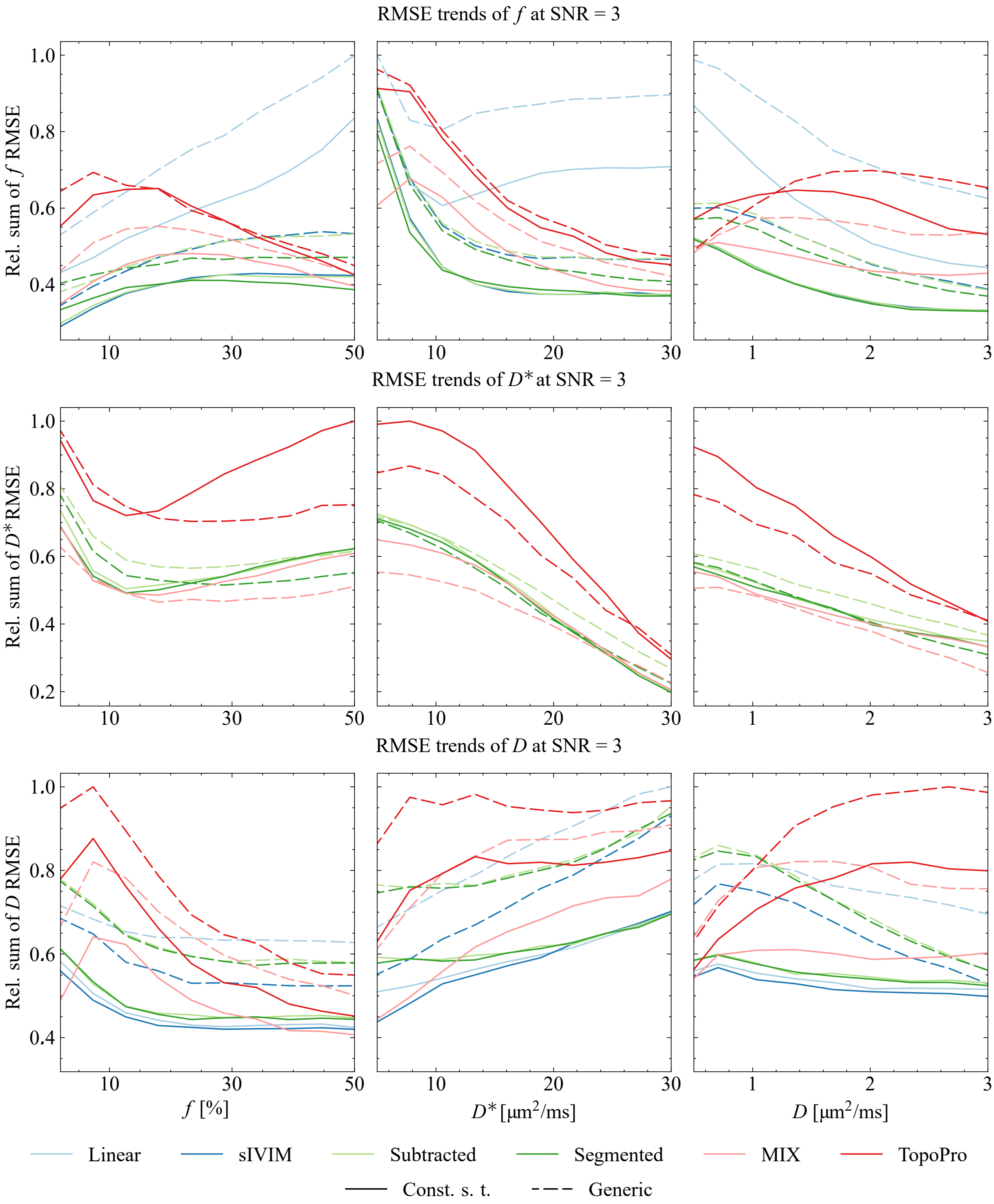


Figure 7: The RMSE trends for SNR 3. Results are only shown for the best and worst performing protocols. All protocol-algorithm combinations can be found in Appendix B.

20%, except the bias which pushed it up towards 30%. MIX did not suffer from any bias and quickly moved towards lower errors with increased SNR. The negative bias of the conventional fits can be seen to arise when D^* approaches D. Similar results can be observed for D in Figure 9, although the sIVIM fit can be seen to perform slightly better than the rest. Errors in D can be expected to range from 0.4 to 0.6 $\mu\mathrm{m}^2\,\mathrm{ms}^{-1}$ at low SNR.

For D^*, Figure 10 shows MIX to be the best estimator. One may note the clearer distinction in these maps, compared to the gap between the curves in Figure 7. Although, looking at the RMSE scale, the errors appear to be in the same range of where one is expected to find D^*, or as many studies in the meta-analysis by He et al. suggest, multiple times larger than the expected values of D^* [26].

For f, Figure 11 shows both a negative and positive bias for the Linear fit, a strictly negative bias for the sIVIM, Subtracted, and Segmented fits, and essentially zero bias for MIX and TopoPro. It is also notable that the *Constant general* protocol showed a slightly lower bias than the rest. It can be seen on the scale that the bias of the conventional fitting algorithms have the ability to range beyond the RMSE of 15-20%. Again, similar results can be seen for D in Figure 12.

The bias in D^* was the lowest for *Constant general* (Figure 13). Similar patterns can be seen for the conventional fits, where D^* became harder to estimate as it approached D. The biases observed for MIX mostly consisted of outliers which indicate that the method is sensitive to the degeneracy of f as it approaches 0. Although, these biases are negligible compared to the RMSE, which can be seen to be an order of magnitude larger.

4.3 Error Dependency on b-values and SNR

Figure 14 shows the contribution of the bias and standard deviation to the total error for different SNR-levels and b-value thresholds. For f and D, the bias can be seen to increase with thresholds, while the standard deviation increases with larger b-values. This results in a minimum for each simulated SNR, which is to be considered as the optimal b-value threshold for that particular b-value. For both f and D, the optimal b-value threshold can be seen to be as low as $150\ \mathrm{s\,mm^{-2}}$ for an SNR-level of 3. For D^*, the bias is negligible in comparison to the standard deviation, making lower thresholds result in lower RMSE.

An additional trend of note is that higher b-values are permitted as SNR increases, as can be seen by the flattening of the surface along the b-axis, compared to the greater curvature seen at the lowest SNR.

4.4 In-vivo Measurements and Evaluation

Figure 15 shows an example of parametric maps of f and D obtained using the MIX algorithm on data provided by the *Generic* protocol. The tumour ROI can be seen to be in agreement with the parametric maps, indicating feasibility of obtaining diffusion and perfusion data from the same measurement.

The observed parameter distributions for f and D^* were skewed, as depicted by the box plots for all protocol-algorithm combinations in Figure 16, with histograms of Segmented and MIX estimates in Figure 17. The effect of protocol optimization can be seen in the box plots of f and D, where a reduced skewness is observed for f, and a negative shift for D, which is the same difference observed between Segmented and MIX in the histograms.

The medians of the box plots in Figure 16 show that f can be expected to be slightly larger (10-12%) in the healthy prostate, compared to the lower values found in the tumour (5-10%). The values are however dependent on both protocol and fitting algorithm. Similar results can be seen for D, where the medians are lower in the tumour compared to the counterparts in the healthy prostate. The D^* distributions of TopoPro (Figure 16) can be seen to exceed the upper bound, with pile-ups at several values of D^*.

Table 4 shows the success ratios of bi-exponential fitting. The conventional methods can be seen to benefit from optimization. While results are more inconclusive for MIX and TopoPro, they do lean towards better performance with the *Generic* protocol.

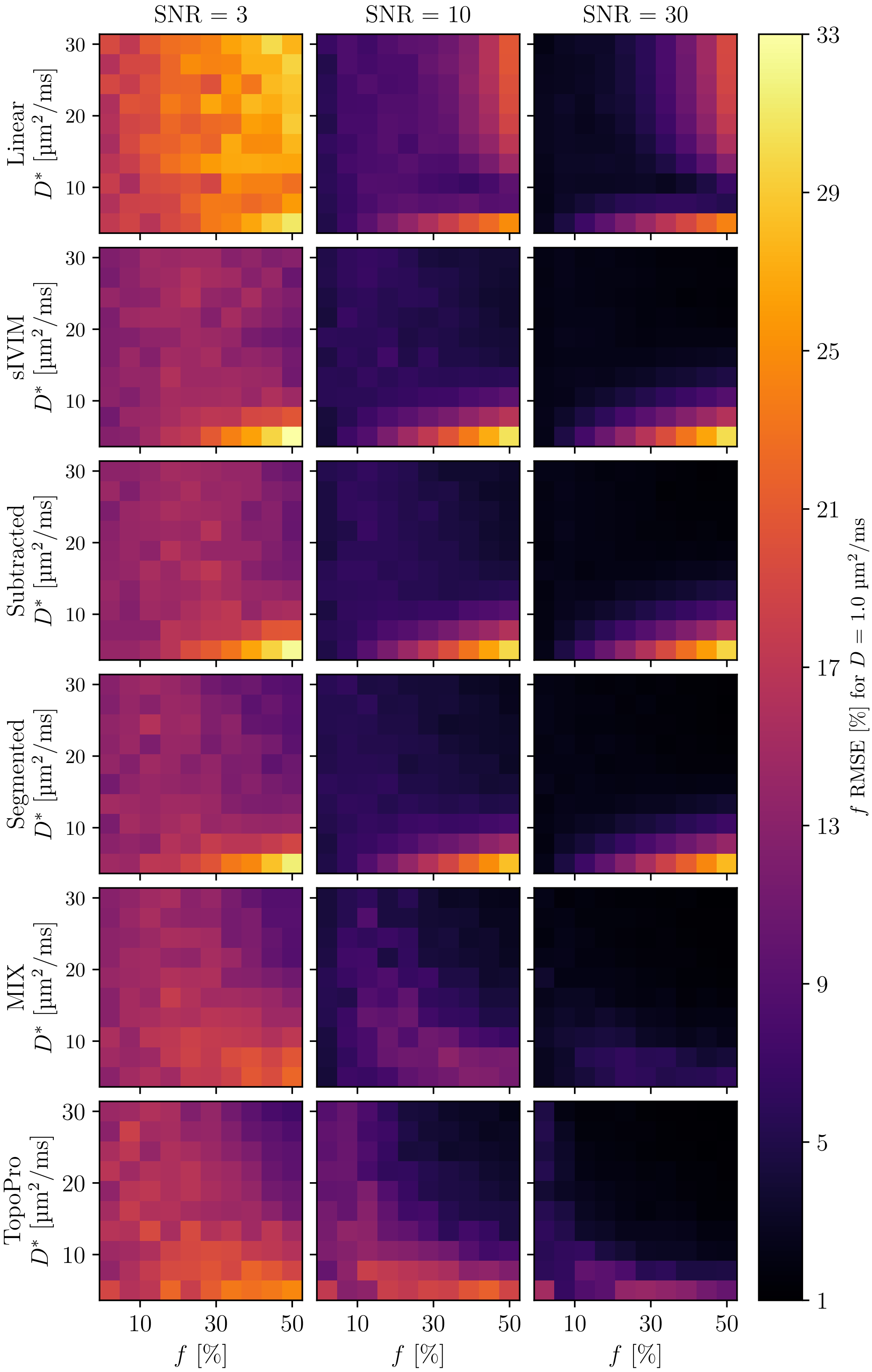


Figure 8: The RMSE of f estimates for different ground truth values of f and D^* with a fixed D of 1.0 µm^2 ms^{-1}. Data simulated using the *Constant single tissue* protocol. Note that the colours are mapped to f in terms of absolute percentage units, and should not be interpreted as relative errors.

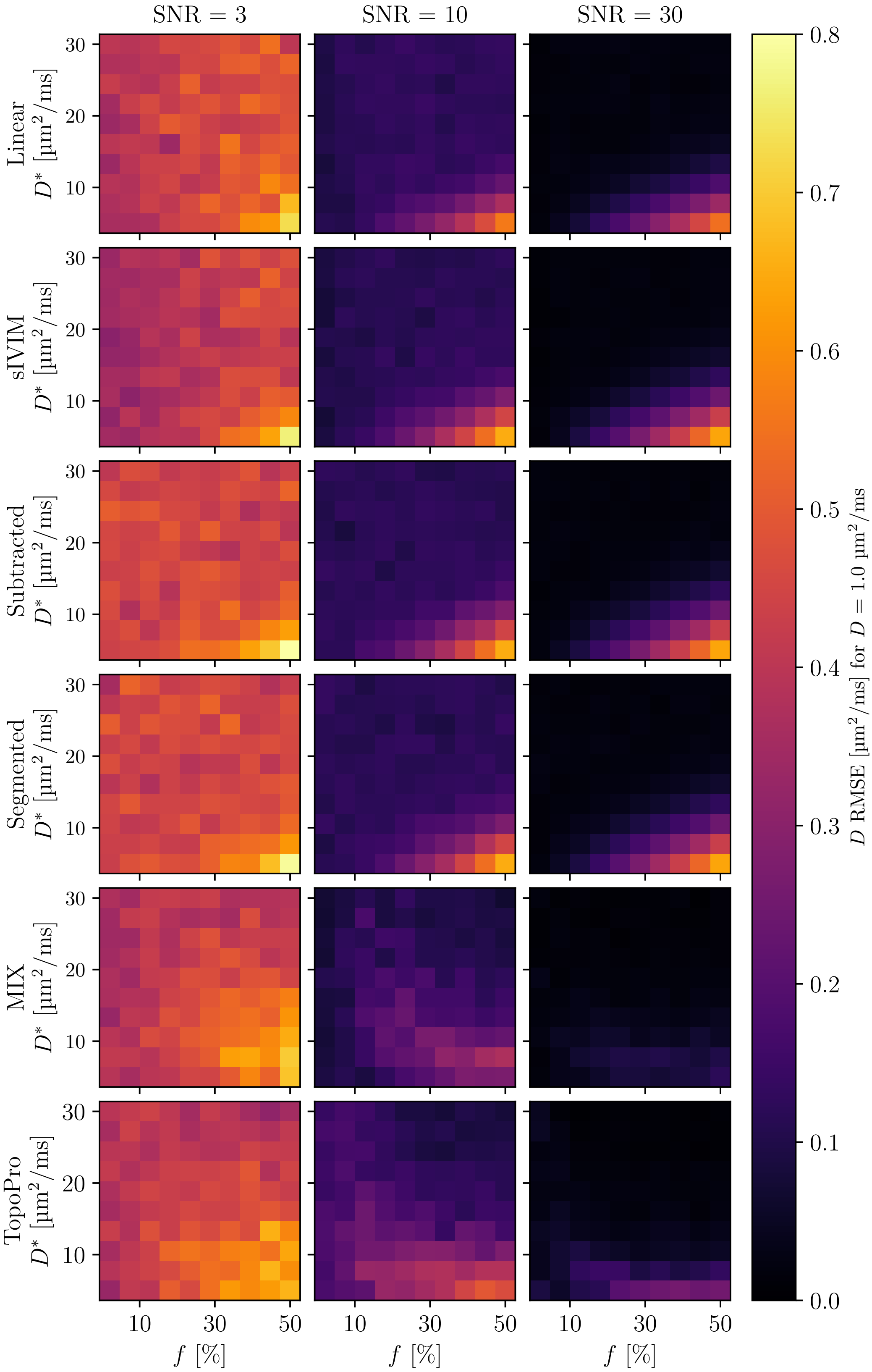


Figure 9: The RMSE of D estimates for different ground truth values of f and D^* with a fixed D of 1.0 μm^2 ms^{-1}. Data simulated using the *Constant single tissue* protocol.

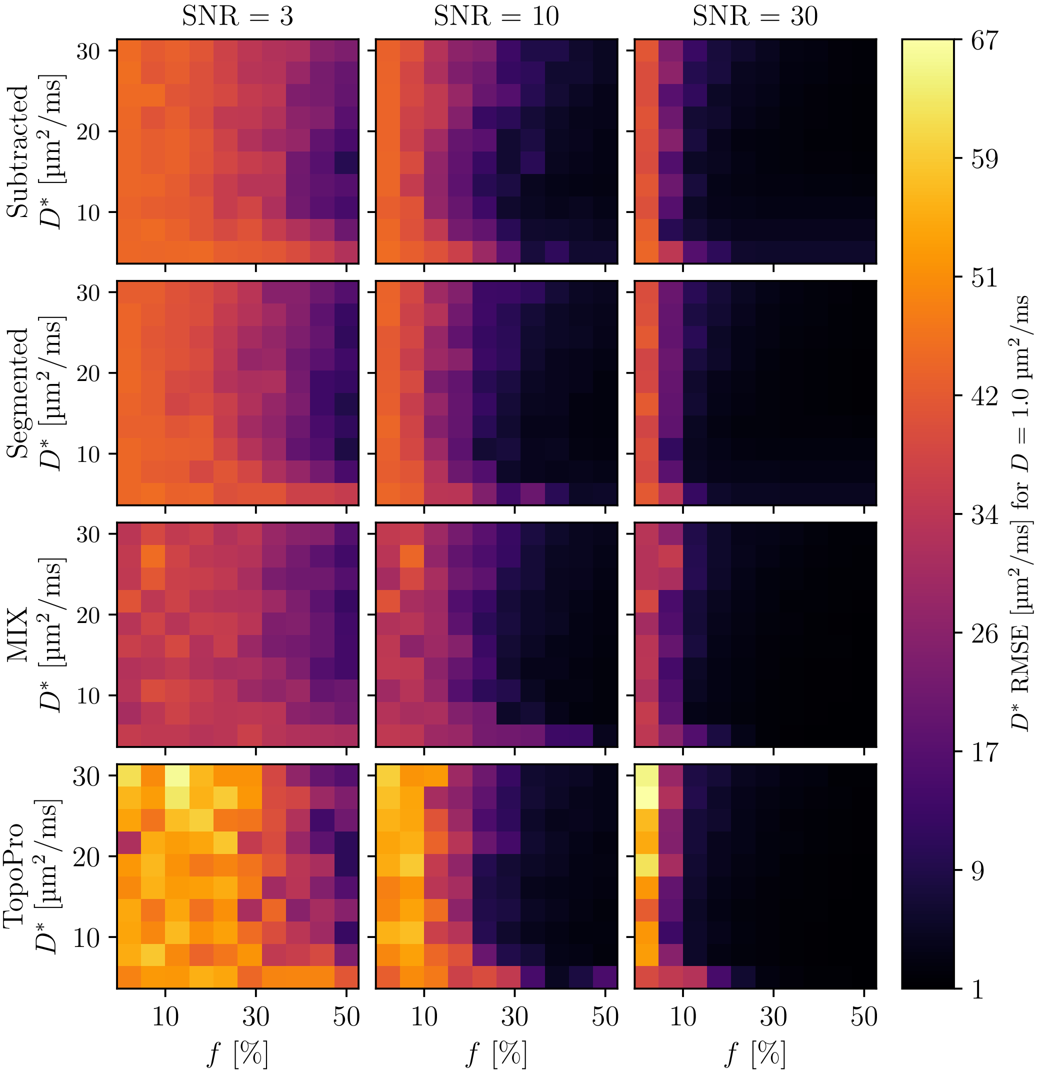


Figure 10: The RMSE of D^* estimates for different ground truth values of f and D^* with a fixed D of 1.0 µm^2 ms^{-1}. Data simulated using the *Constant single tissue* protocol. Note that only four of the six fitting methods are able to provide estimates for D^*.

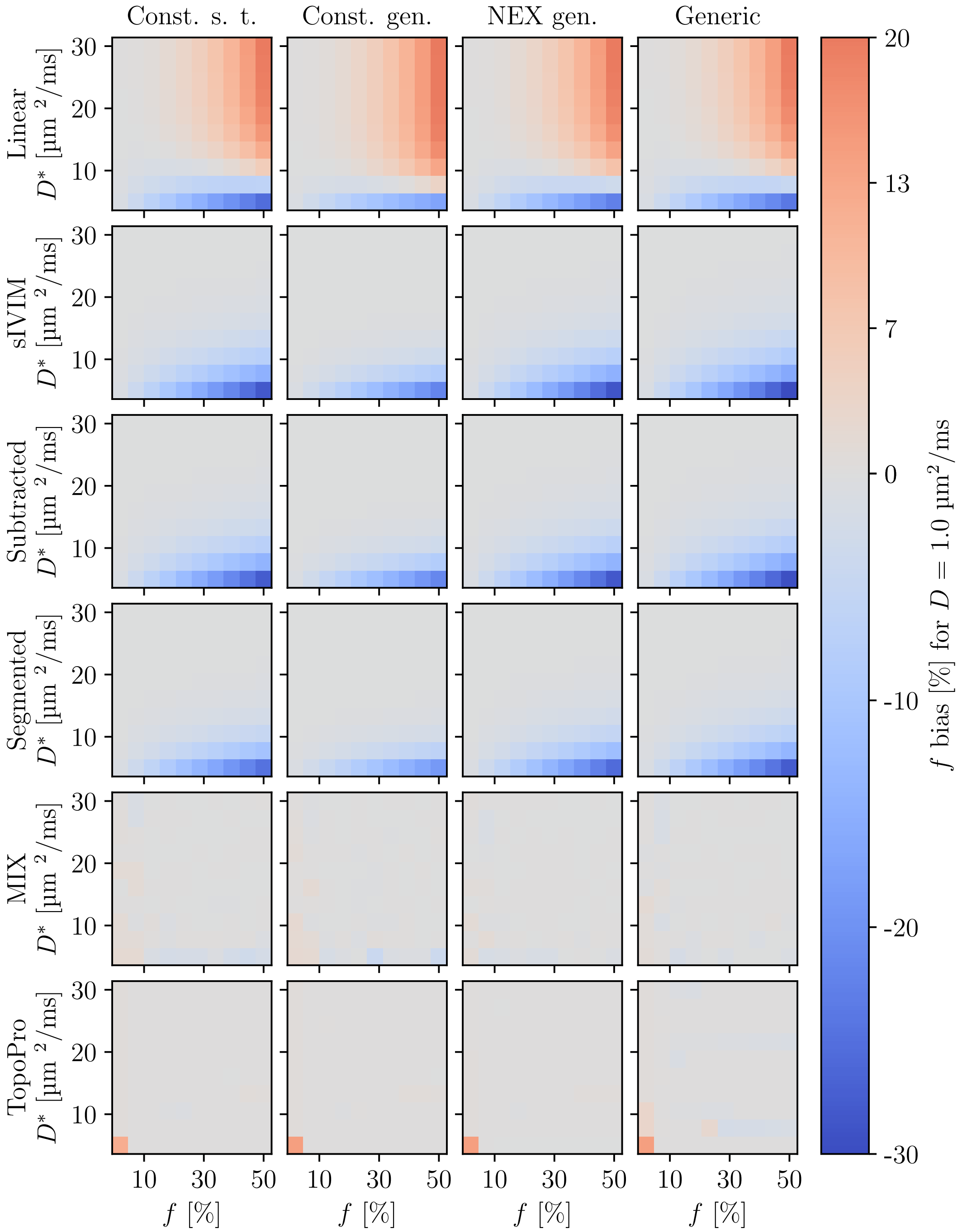


Figure 11: The bias in f estimates for different ground truth values of f and D^* with a fixed D of 1.0 $\mu m^2\,ms^{-1}$. Note the different scales for positive and negative values. Also note that the colours are mapped to f in terms of absolute percentage units, and should not be interpreted as relative bias.

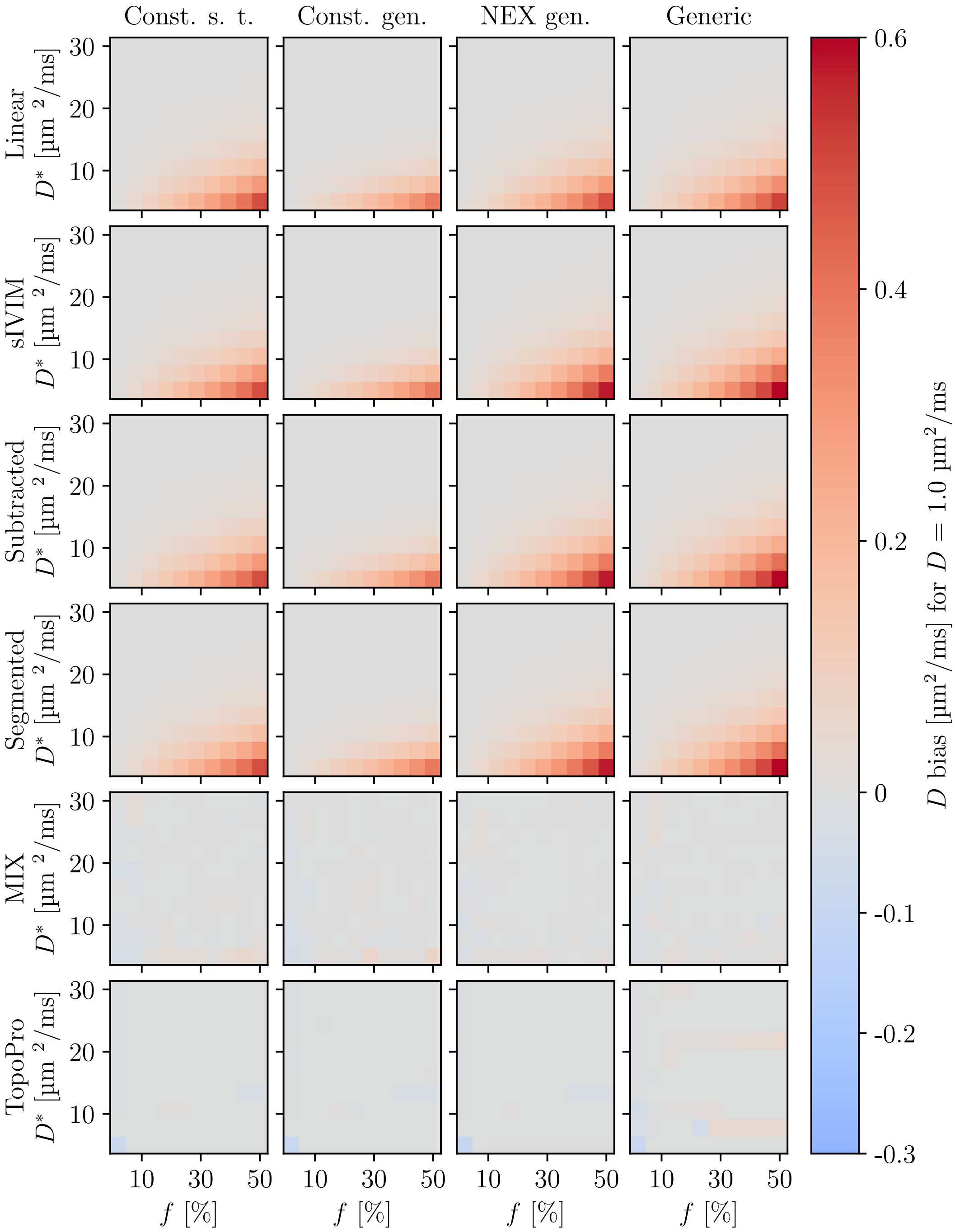


Figure 12: The bias in D estimates for different ground truth values of f and D^* with a fixed D of 1.0 $\mu m^2\,ms^{-1}$. Note the different scales for positive and negative values.

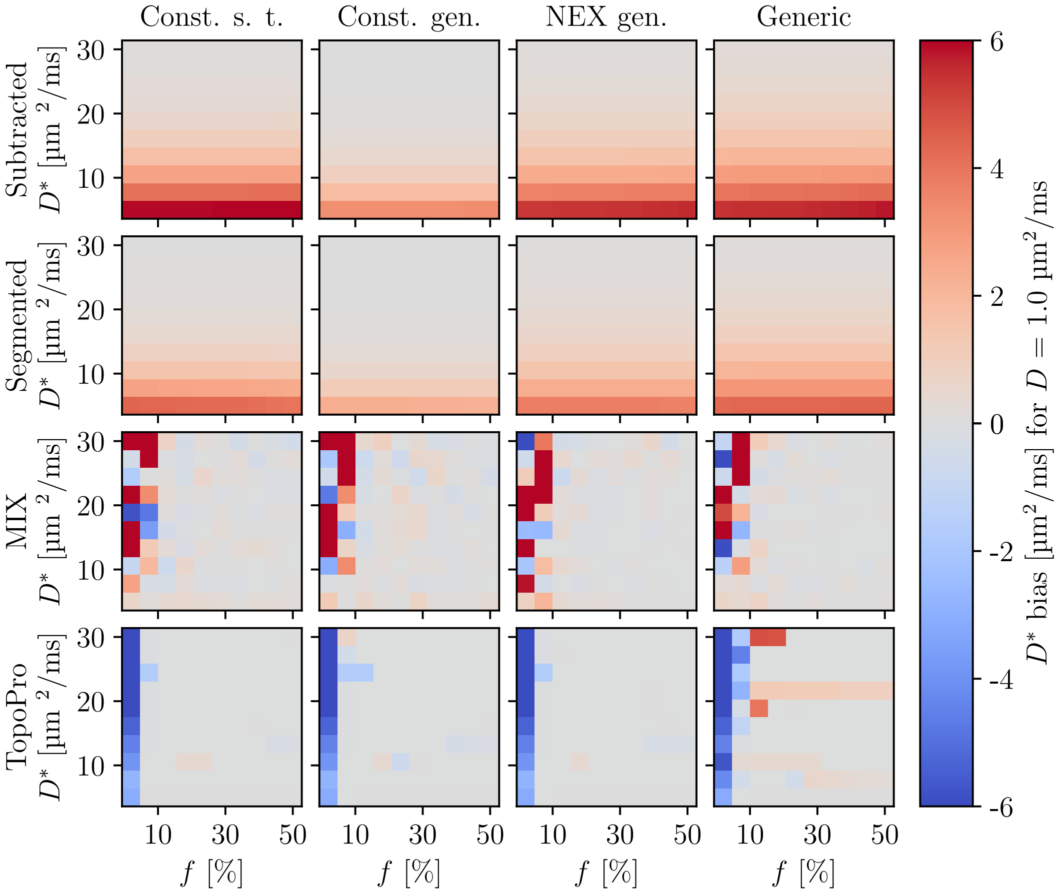


Figure 13: The bias in D^* estimates for different ground truth values of f and D^* with a fixed D of 1.0 μm^2 ms^{-1}. Note the different scales for positive and negative values.

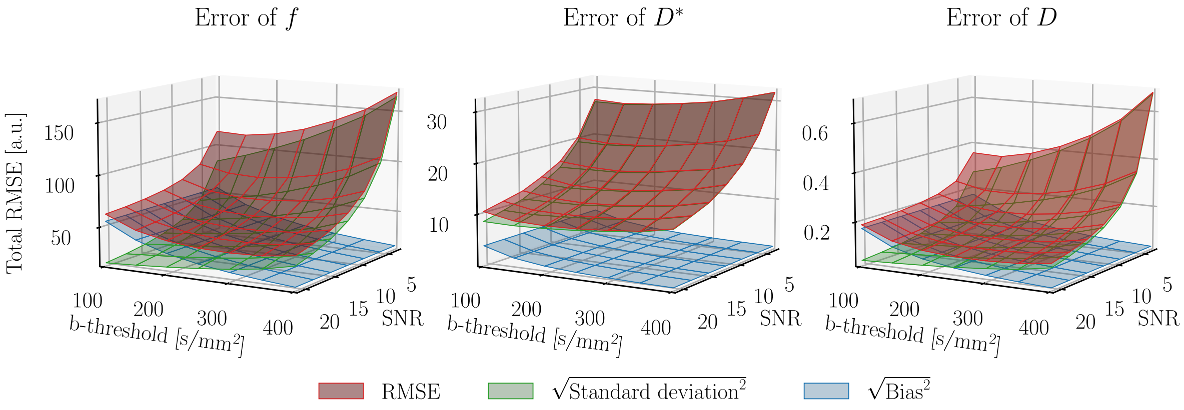


Figure 14: The RMSE and its two contributors as a function of SNR and b-value. Note that the green standard deviation surface in the plot of D^* error is just under the RMSE surface, and is thus not visible.

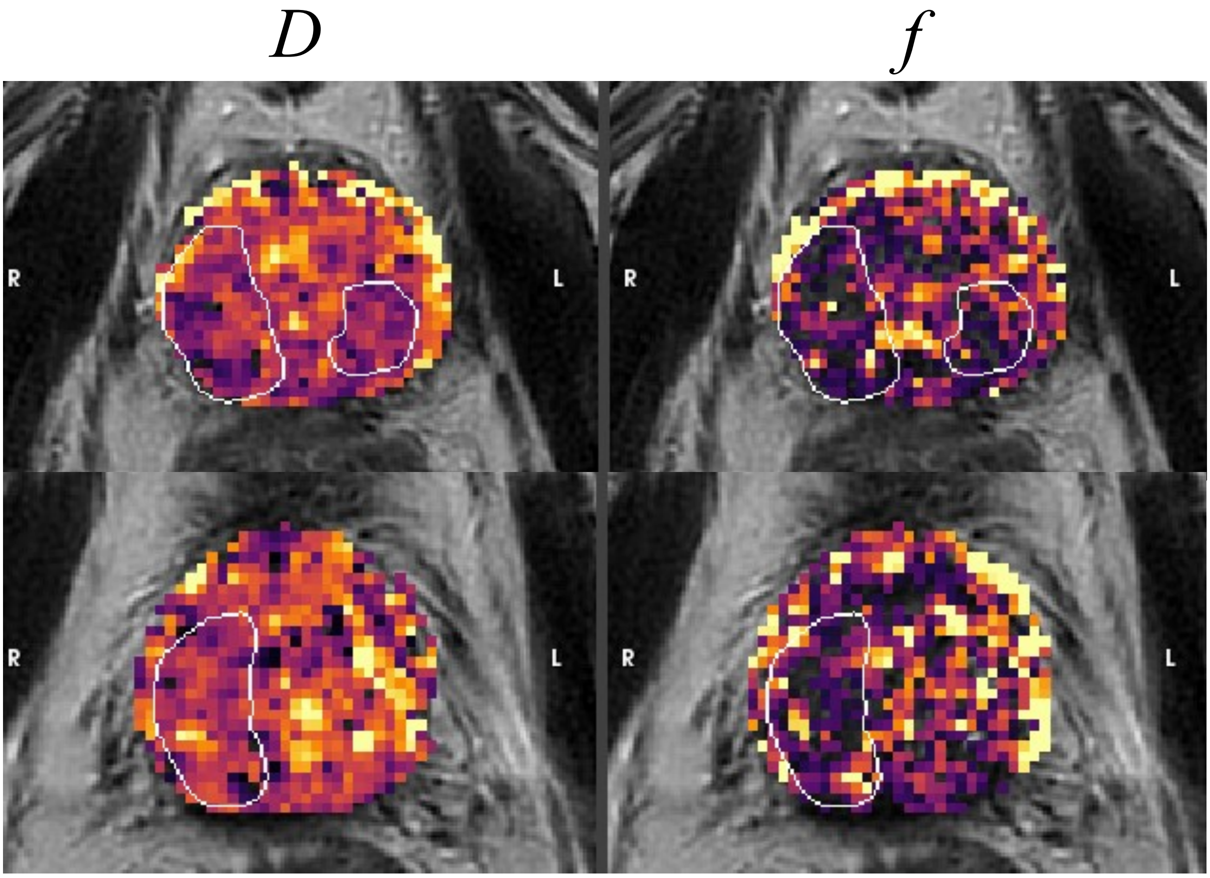


Figure 15: Parametric maps of D and f with tumour delineations by an oncologist. Two different slices of the same patient. The ROI can be seen to correspond with regions that show lower values of D and f. This indicates higher cell density and lower perfusion respectively.

Table 4: The success ratios of f (upper table) and D^* (lower table) for each patient and fitting algorithm. Note that > 1 indicates that the optimized protocol provides more successful bi-exponential fits, vice versa for < 1. One standard deviation is reported with the averages.

$\boldsymbol{f}$	Linear	sIVIM	Subtracted	Segmented	MIX	TopoPro
Patient 1	1.20	1.19	1.19	1.12	1.03	1.01
Patient 2	1.05	1.02	1.04	1.00	0.96	0.96
Patient 3	1.08	1.11	1.10	1.07	1.02	1.01
Patient 4	0.94	1.06	1.04	0.99	0.87	0.88
Patient 5	0.97	1.06	1.05	1.01	0.90	0.90
Average	**1.05(9)**	**1.09(6)**	**1.08(6)**	**1.04(5)**	**0.96(6)**	**0.95(5)**
$\boldsymbol{D^*}$						
Patient 1			1.18	1.14	0.94	0.95
Patient 2			1.11	1.06	0.89	0.94
Patient 3			1.09	1.07	0.95	1.00
Patient 4			1.13	1.06	1.18	0.85
Patient 5			1.05	0.98	1.05	0.91
Average			**1.11(4)**	**1.06(5)**	**1.00(10)**	**0.93(5)**

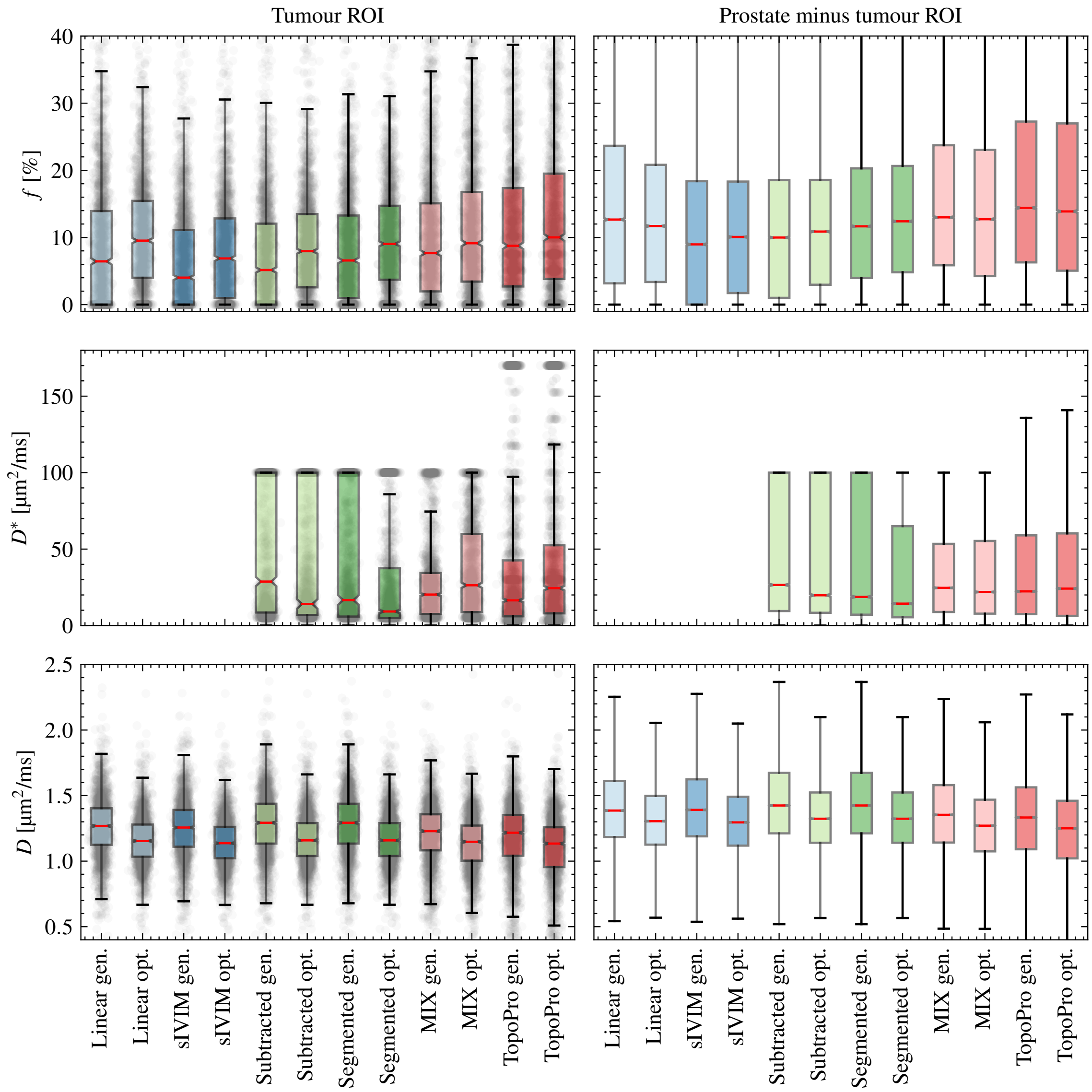


Figure 16: Box plots of parameter distributions acquired for different algorithm-protocol combinations. Note that the Linear and sIVIM algorithms do not provide estimates of D^*. Data points are jittered under the tumour ROI boxes to show the underlying distribution. Note the pile-up at the bounds of the f and D^* distributions, these are voxels with data points that failed to depict bi-exponential signal behaviour.

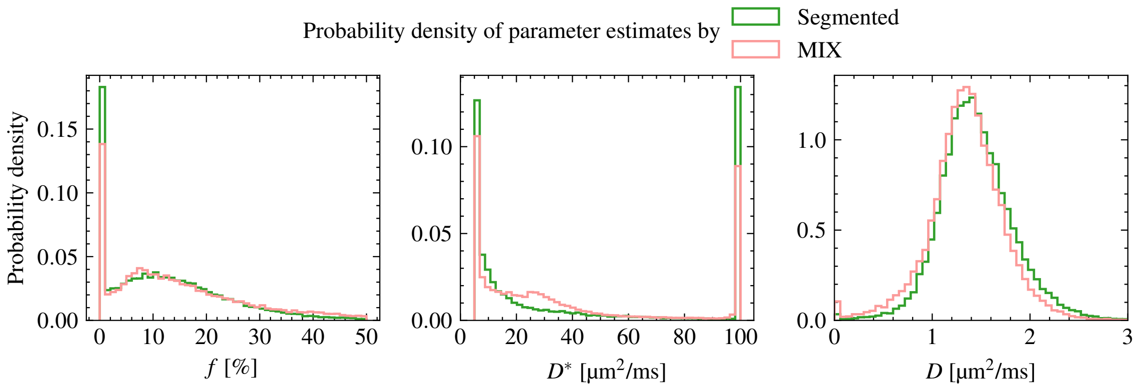


Figure 17: Probability density functions of parameter estimates provided by the MIX and Segmented algorithms on data acquired with the generic protocol.

5 Discussion

5.1 Evaluation of Fitting Algorithms

The results of the RMSE trends and RMSE maps show that the conventional sIVIM, Subtracted, and Segmented methods perform well regarding noise sensitivity. However, they do show a bias as D^* approaches D, therefore making them less ideal for usage in tissues that are known to have low D^*. In such cases, bias free methods such as MIX and TopoPro should be looked into.

The Linear fit showed profoundly greater RMSE in f compared to the other methods, while the comparable sIVIM did not. This is due to the Linear fit relying on the SNR of b-values greater than the b-value threshold, where SNR is lower, making it sensitive to noise. The sIVIM fit on the other hand handles noise well. This is due to b-value $0\ \mathrm{s\,mm^{-2}}$ being included in the fit, which contributes with better SNR, thus more certain estimation of f. It is therefore recommended that sIVIM is used in any instance where the Linear fit is intended to be used, as it is a fast method that similar to the Linear fit, only requires three b-values. Furthermore, sIVIM is a good alternative to the slower Subtracted and Segmented methods, as the latter two require more non-linear least-squares fit to be performed, while sIVIM only requires a single fit. However, this is only if D^* is not of interest, as sIVIM is unable to estimate it.

Looking at the bias maps (Figures 11-13) for the conventional methods, the bias can be seen to vary more depending on the protocol rather than fitting method, especially in Figure 13, where *Constant general* shows the lowest bias. This is due to two of the protocols having different "effective" b-value thresholds in the fits, as they do not all sample b-value $200\ \mathrm{s\,mm^{-2}}$. Interestingly, although the *Constant general* showed worse RMSE (Appendix B) than *Constant single tissue*, its bias was lower. It was this revelation that prompted the investigation into the b-value threshold for the segmentation-based fitting methods and its impact on the RMSE.

MIX and TopoPro show no bias, making them better at higher SNR where the other methods show a bias, especially if D^* is known to be low in the tissue of interest. Figure 7 show that MIX has lower RMSE than TopoPro at low SNR. Furthermore, MIX shows to be the best estimator of D^* in Figure 10. MIX is therefore a well performing bias-free alternative to the conventional methods. However, with the current implementations, the computation time of both MIX and TopoPro were approximately 15 times longer, which is worthy of consideration if time is a factor.

Another point of note regarding MIX and TopoPro, they both show less variations in performance between the different protocols (Appendix B). The gaps between the best and worst protocols are generally narrower than the ones observed for the conventional methods. Therefore, MIX and TopoPro are candidates for fitting algorithms which could be used to compare results of different studies where b-value schemes are known to vary. Furthermore, MIX and TopoPro do not depend on any b-value thresholds, which is another variable that can differ between studies. This can further be seen in the box plots of Figure 16, where the difference in medians between the two protocols is smaller for MIX and TopoPro, compared to the other methods.

In-vivo, MIX was the only method that provided meaningful estimates of D^*, as seen in Figure 17, where a distribution can be observed, in contrast to the regularization pattern seen in the distribution from the Segmented method. MIX further provides more bi-exponential fits, as can be seen by the smaller pile-up at $f = 0\%$. Using the results of the simulations as reference, this pile-up could be caused by the negative bias seen for the conventional methods (Figure 11). The algorithm essentially fails to identify the bi-exponential behaviour of the signal when D^* is low, hence the mono-exponential fit by setting f to zero. This is a relevant issue in the prostate, as many studies included in the meta-analysis by He et al. [26] reported D^* values less than $10\ \mathrm{\mu m^2\,ms^{-1}}$. It is indeed in this range of D^*-values where degeneracy occurs, as seen in the bias maps of Figure 11.

The differences in the distribution of D estimates manifest as a shift, where MIX shows lower estimates than the Segmented method. This is due to D being overestimated when $f = 0\%$, as only a mono-exponential fit is performed. These mono-exponential fits are less frequent when MIX is used, and should therefore be regarded as the more truthful distribution compared to the Segmented method.

One may note that it was indeed this overestimation of D that lead to the bi-exponential IVIM model, as discussed in Section 2.2.

Histograms for the rest of the conventional algorithms were not shown as the distribution of the Segmented method showed that it performed the best out of the four. Regarding TopoPro, it did show slightly more successful bi-exponential fits. However, D^* was heavily regularized and showed unreliable pile-ups across the entire scale, likely caused by dynamic bound setting in its third optimization step. It should be recognized that TopoPro could still be improved by fine-tuning the bounds, which was not done in this work. One could argue that our results, shown and not shown, indicate that MIX works better straight out of the box, while TopoPro would need more modifications to it in order to provide reliable estimates.

Pooling the results of the simulations and in-vivo evaluation, it can be concluded that MIX may be the superior fitting algorithm, as it is able to identify the bi-exponential signal more frequently compared to the other methods, and was the only algorithm able to provide a parameter distribution for D^*. The clinical implication this carries is that the reliability of IVIM-derived perfusion and diffusion data can be increased by simply changing fitting algorithm. Furthermore, our evaluation showed that the Linear fitting algorithm which has currently been used in creation of hypoxia maps was the algorithm that showed the lowest performance. Future IVIM-derived hypoxia maps can therefore be improved, which could potentially yield even more promising results and strengthen their candidacy for standardized hypoxia mapping of the prostate.

5.2 Error Dependency on b-value Threshold and SNR for Segmentation-based Methods

The optimal b-value threshold of $150\,\mathrm{s\,mm^{-2}}$ found for low SNR is lower than the usually recommended interval of 200 to $400\,\mathrm{s\,mm^{-2}}$ [10], where $200\,\mathrm{s\,mm^{-2}}$ is most commonly used [46]. However, our results are in agreement with Wurnig et al. [33], who also found that lower thresholds performed better when evaluated in-vivo. They found that a threshold as low as $20\,\mathrm{s\,mm^{-2}}$ could be the most optimal for well perfused organs such as the liver, while as in our case, $150\,\mathrm{s\,mm^{-2}}$ proved optimal for e.g. the kidney. A limitation of our evaluation is that it was general, and incorporated the errors for a wide range of tissue properties. Previous knowledge, and indeed the results of Wurnig et al., show that optimal b-value thresholds need to be found for the organ of interest, as general recommendations such as $200\,\mathrm{s\,mm^{-2}}$, or even $150\,\mathrm{s\,mm^{-2}}$ may not result in the best possible parameter estimation.

Another insight provided by this evaluation of b-value thresholds is the error contributions of the bias and noise sensitivity of the conventional fitting algorithms. As the b-value threshold is lowered, data points with better SNR are included in the estimation of D, making its estimation more certain. Thus, lower errors are propagated to f and D^*, as D is fixed in their estimation. However, as the lowering of the threshold may decrease noise sensitivity, the bias in increased. Therefore, one sacrifices accuracy in the pursuit of better precision. At some point, the best accuracy and best precision is achieved, such that their combination; the RMSE; is minimized. This point is the optimal b-value threshold, which is seen in Figure 14 to be dependent on SNR.

As discussed earlier, a side effect of only sampling four b-values is that the sparsity may lead to effective b-value thresholds, as even if the threshold is set to e.g. $200\,\mathrm{s\,mm^{-2}}$, if only b-values $350\,\mathrm{s\,mm^{-2}}$ and $800\,\mathrm{s\,mm^{-2}}$, exist in that range, the threshold for the fit effectively becomes $350\,\mathrm{s\,mm^{-2}}$. This is the case for the *Constant general* protocol, which showed the lowest bias. This tells us that b-value optimization needs to account for the bias of the fitting algorithm that will be used to analyse the data. This is currently not accounted for in Cramér-Rao optimizations, whereas optimizations based on estimation uncertainty do, as done by e.g. Paganelli et al. for the Linear fit [47].

5.3 b-value Optimization and Protocol Finalization

In this work, as Table 1 shows, we have compared a diverse b-value set optimized for several types of tissue properties (*NEX general*) to one typically seen in literature (*Generic*), where the optimized protocol showed better performance (Appendix B). Although, *NEX general* only performed marginally better than

Generic. We also compared two protocols with fewer b-values and higher SNR, where the protocol that was optimized for a specific IVIM parameter set; and can also be characterized by its lower b-values, thus even better SNR; showed better performance than a protocol optimized for a wide range of tissues. Ultimately, the conclusion of Table 3 is that any Cramér-Rao optimized protocol performed better than the *Generic*.

In-vivo, the reduced skewness of the f box plots in Figure 16 show that by improving the acquisition, the mono-exponential f estimates of 0% can be reduced. The histograms of Figure 17 show that the skewness is mainly caused by mono-exponential fits. Hence the reduced skewness of the optimized acquisitions may be due to more bi-exponential fits. Although the quantitative measure of the successful bi-exponential fits proved this for the conventional methods, it did not for MIX. This could be attributed to the averaging of the histograms, as the values in Table 4 can be seen to not only depend on the algorithm used, but also on the patient imaged. Hence, effects of a single patient may affect the entire distribution.

The reason for employing histogram averaging was due to the small sample size. Ideally, the differences in medians between tumour and healthy tissue for each patient and protocol would have been determined. Using such differences would have further allowed for proper statistical testing and eliminated confounding effects caused by e.g. prostate and tumour size (affecting the number of voxels in the ROI, hence the distributions). Our sample size of 5 patients was deemed too small to draw any certain conclusions from such an analysis.

Regarding TopoPro, Figure 16 shows clusters of estimates at multiple values of D^*. These are caused by dynamic bound setting, as the bounds in the third step of the TopoPro algorithm are based on parameter estimates from the first and second steps. Hence, failed estimates are piled up on the lower and upper edges of these new bounds, instead of the typical values seen for the other algortihms. Therefore, these should not be interpreted as peaks of distributions that reflect tissue properties. Furthermore, because of these clusters of estimates, our definition of successful bi-exponential fits for D^* becomes invalid, hence invalidating the results of the D^* evaluation shown in Table 4. For these reasons, TopoPro-derived results of protocol evaluation in-vivo should be interpreted with caution, as the results may greatly depend on the settings of the algorithm.

A further point of note regarding the quantifications of successful bi-exponential fits presented in Table 4. These values only compare the two protocols for a given patient and algorithm. They cannot be used for comparisons between algorithms as the magnitude of the quantity measures the performance increase or decrease for a particular algorithm given two protocols, i.e. protocol dependency in performance. If an algorithm shows greater performance with one or the other protocol, it does not mean that it performs better than an algorithm that is less dependent on the protocol that is being used.

Further limitations of the in-vivo tumour versus healthy prostate comparison are that the tumour ROI were drawn by an oncologist and not a radiologist, which would have been more appropriate. Furthermore, the tumours were delineated on T2-weighted images acquired for the purposes of radiotherapy treatment planning, and were acquired in the same examination as the diffusion acquisition for the IVIM analysis. These T2-weighted images are not of diagnostic quality as their purpose is to delineate the outer contours of the prostate. These T2-weighted images have worse contrast within the prostate when compared to diagnostic quality T2-weighted images. Therefore, discrepancies were expected. Ideally, diagnostic quality T2-weighted images should have been acquired during the same exam as the IVIM diffusion-weighted images. Our ethical permission set limits on how much measurement time we could add to an exam, which in our case, was used for the IVIM acquisition.

There are also a number of limitations in our simulations and optimizations. The first being that the parameter values used for the Cramér-Rao optimization for the *Constant single tissue* protocol are not certain to be optimal as the ranges presented in the meta-analysis by He et al. [26] were large. Therefore, *Constant single tissue* may not be fully optimized for the prostate. Furthermore, the simulations only accounted for Rician noise and not physiological noise. Such simulations would require a more intricate virtual phantom.

As discussed earlier, the Cramér-Rao lower bound does not account for estimation bias. Protocols could be optimized further by including these effects. However, our simulation results show that all three

Cramér-Rao-based protocols performed better than the *Generic*, in agreement with previous studies [29, 30, 40]. Some of these studies were able to show a lowered standard deviation of prameter estimates in-vivo, either by ROI analysis, or by voxel-wise analysis after repeated measurements. We were not able to evaluate our in-vivo data in the same way due to known confounding variables or methodological limitations.

Regarding ROI analysis, some tumours were too small, thus consisting of too few voxels to draw certain conclusions from differences seen between the two protocols. This issue is further amplified by the relatively small size of the prostate, where resolution is often limited to increase SNR. Furthermore, there were the previously discussed discrepancies in the tumour ROI delineations. Ideally, since the purpose of this work is to optimize and evaluate IVIM for hypoxia maps, voxel-wise error quantification would have been more useful than ROI analysis, as performed by Jalnefjord et al. [29]. However, this would have required with repeated acquisitions with the same protocol. Due to time constraints in the examination of the patients, we were not able to do this. The allocated time was used to acquire images using two different protocols. Therefore, it would have been more suitable to do such an experiment on healthy volunteers. Despite these methodological limitations, we were able to identify a key issue that has to be resolved with experimental changes. Namely, to reduce the number of mono-exponential fits.

A point of note is that in our in-vivo evaluation, the GE image reconstruction software AIR Recon DL was used. Figure 4 showed that it increased SNR from 3 to 6 at b-value $800\ \mathrm{s\,mm^{-2}}$ for a single excitation of a spherical water phantom. Assuming the same SNR is achieved in-vivo, Figure 14 shows that the optimal b-value threshold for f and D can be seen to be in the range 200-250 $\mathrm{s\,mm^{-2}}$. Table 2 tells us that it is indeed in this range the second highest b-value is located, thus making the effective b-value threshold $240\ \mathrm{s\,mm^{-2}}$. The properties of the protocol therefore theoretically minimized the RMSE of the Segmented method. Had AIR Recon DL not been used, this would have not happened, as the optimal b-value threshold would have been lower than what was sampled.

The ultimate conclusion of the simulations is that protocols with fewer and lower b-values with better SNR perform better than any other constellation of number of b-values and SNR that was tested in this work. The in-vivo evaluation of the optimized protocol was more inconclusive due to a small sample size, but did show increased ability to fit the bi-exponential IVIM model to data for the Linear, sIVIM, Subtracted, and Segmented algorithms.

5.4 Future Prospects

In this work, the diffusive IVIM model of eq. (2) was studied. This is a very basic study, useful in learning the limitations and problems of IVIM. There are a number of future paths where different models and experimental methods can explore to further increase the reliability of IVIM parameter estimates. Furthermore, there are a number of methods that have been applied to other organs such as the brain and the liver, which has not yet had their usefulness evaluated on the prostate.

The diffusive IVIM model assumes that the blood flow changes direction several times during the diffusion encoding time. As mentioned in Section 2.3, this assumption has been violated in other organs and to our knowledge, has not been evaluated for the prostate. Flow-compensated measurements with short diffusion times are therefore warranted to probe the ballistic limit. Depending on the prostates vascular characteristics, such experiments may require MR-systems with strong magnetic field gradients.

Accounting for partial volume effects has shown to improve results in the brain by considering T2 relaxation [48, 49]. It remains to be seen how partial volume effects affect the results in prostate IVIM. Chatterjee et al. [50, 51] have shown how a combination of diffusion and T2 properties can be used to perform tissue decomposition. Furthermore, there are partial volume effects in the diffusion coefficient caused by different types of tissue microstructure. These can be separated using tensor-valued diffusion encoding [15]. The usefulness of this in combination with IVIM remains to be evaluated for the prostate.

Regarding the topics of this work, optimization of data points and parameter estimation; these issues will have to be revisited for emerging experimental techniques that show promise for prostate applications. A relevant example is a recent study on optimizing flow-compensated abdominal measurements [23]. Consensus regarding optimization method is yet to be reached. New and current fitting algorithms should continually be evaluated and standardized, which is the objective of the newly formed IVIM task force within the Open Science Initiative for Perfusion Imaging project [52].

6 Conclusions

Our evaluation on both simulated and in-vivo data has shown that the reliability of IVIM-derived perfusion and diffusion parameters can be improved by optimizing acquisition protocols, and by changing fitting algorithms and their settings. These results can be used to optimize acquisitions and improve analysis for the purposes of non-invasive hypoxia mapping of the prostate, and IVIM in general.

We showed via simulations that three of the four conventional fitting algorithms handle noise well, but with a negative bias for low D^*. The Linear fit is very sensitive to noise. We instead recommend the use of sIVIM, as it is a more robust alternative that similar to the Linear fit, only requires three b-values. Out of the two new methods, MIX is an accurate and precise algorithm that excelled in the in-vivo evaluation, but does however have an approximately 15 times longer computation time.

We recommend usage of lower b-value threshold in the Linear, sIVIM, Subtracted and Segmented algorithms, as we found it to reduce noise sensitivity. A general recommendation would be $150\,\mathrm{s\,mm^{-2}}$.

Cramér-Rao optimization of b-values has in simulations shown to produce more robust protocols than typical ones found in literature. We showed that fewer and lower b-values with higher SNR perform better than many b-values with lower SNR. The use of such protocols increased the in-vivo performance of the Linear, sIVIM, Subtracted and Segmented algorithms.

References

1. Björk-Eriksson T. Strålbehandling – Behandling och biverkningar. Cancerfonden. Available from: `https://www.cancerfonden.se/om-cancer/behandlingar/stralbehandling` [Accessed on: 2023 Apr 20]
2. Bratt O. Prostatacancer – Symtom, prognos och behandling. Cancerfonden. Available from: `https://www.cancerfonden.se/om-cancer/cancersjukdomar/prostatacancer` [Accessed on: 2023 Apr 20]
3. Minniti G, Goldsmith C and Brada M. Chapter 16 - Radiotherapy. In: *Handbook of Clinical Neurology*. Ed. by Aminoff MJ, Boller F and Swaab DF. Vol. 104. Elsevier, 2012 Jan 1:215–28
4. Olsson LE, Johansson M, Zackrisson B and Blomqvist LK. Basic Concepts and Applications of Functional Magnetic Resonance Imaging for Radiotherapy of Prostate Cancer. Phys Imaging Radiat Oncol. 2019 Jan 1; 9:50–7
5. Gérard M, Corroyer-Dulmont A, Lesueur P, Collet S, Chérel M, Bourgeois M et al. Hypoxia Imaging and Adaptive Radiotherapy: A State-of-the-Art Approach in the Management of Glioma. Front Med. 2019; 6:117
6. Hompland T, Hole KH, Ragnum HB, Aarnes EK, Vlatkovic L, Lie AK et al. Combined MR Imaging of Oxygen Consumption and Supply Reveals Tumor Hypoxia and Aggressiveness in Prostate Cancer Patients. Cancer Res. 2018 Aug 15; 78(16):4774–85
7. Iima M. Perfusion-Driven Intravoxel Incoherent Motion (IVIM) MRI in Oncology: Applications, Challenges, and Future Trends. Magn Reson Med Sci. 2021 Jun 1; 20(2):125–38
8. Federau C. Intravoxel Incoherent Motion MRI as a Means to Measure in Vivo Perfusion: A Review of the Evidence. NMR Biomed. 2017; 30(11):e3780
9. Koh DM, Collins DJ and Orton MR. Intravoxel Incoherent Motion in Body Diffusion-Weighted MRI: Reality and Challenges. AJR Am J Roentgenol. 2011 Jun; 196(6):1351–61
10. Le Bihan D. What Can We See with IVIM MRI? Neuroimage. 2019 Feb 15; 187:56–67
11. Li YT, Cercueil JP, Yuan J, Chen W, Loffroy R and Wáng YXJ. Liver Intravoxel Incoherent Motion (IVIM) Magnetic Resonance Imaging: A Comprehensive Review of Published Data on Normal Values and Applications for Fibrosis and Tumor Evaluation. Quant Imaging Med Surg. 2017 Feb; 7(1):598–78
12. Paschoal AM, Leoni RF, dos Santos AC and Paiva FF. Intravoxel Incoherent Motion MRI in Neurological and Cerebrovascular Diseases. Neuroimage: Clin. 2018 Jan 1; 20:705–14
13. Szubert-Franczak AE, Naduk-Ostrowska M, Pasicz K, Podgórska J, Skrzyński W and Cieszanowski A. Intravoxel Incoherent Motion Magnetic Resonance Imaging: Basic Principles and Clinical Applications. Pol J Radiol. 2020; 85(1):624–35
14. Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E and Laval-Jeantet M. MR Imaging of Intravoxel Incoherent Motions: Application to Diffusion and Perfusion in Neurologic Disorders. Radiol. 1986 Nov; 161(2):401–7
15. Szczepankiewicz F, Westin CF and Nilsson M. Gradient Waveform Design for Tensor-Valued Encoding in Diffusion MRI. J Neurosci Methods. 2021 Jan 15; 348:109007
16. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J and Laval-Jeantet M. Separation of Diffusion and Perfusion in Intravoxel Incoherent Motion MR Imaging. Radiol. 1988 Aug; 168(2):497–505
17. Federau C. Measuring Perfusion: Intravoxel Incoherent Motion MR Imaging. Magn Reson Imaging Clin N Am. 2021 May 1; 29(2):233–42
18. Wetscherek A, Stieltjes B and Laun FB. Flow-Compensated Intravoxel Incoherent Motion Diffusion Imaging. Magn Reson Med. 2015; 74(2):410–9
19. Jalnefjord O, Geades N and Ljungberg M. On the Bias Introduced by Assuming Different Intravoxel Incoherent Motion Regimes. In: *Proceedings of the 28th Annual Meeting of ISMRM*. 2020 :4434
20. Kennan RP, Gao JH, Zhong J and Gore JC. A General Model of Microcirculatory Blood Flow Effects in Gradient Sensitized MRI. Med Phys. 1994 Apr; 21(4):539–45
21. Ahlgren A, Knutsson L, Wirestam R, Nilsson M, Ståhlberg F, Topgaard D et al. Quantification of Microcirculatory Parameters by Joint Analysis of Flow-Compensated and Non-Flow-Compensated Intravoxel Incoherent Motion (IVIM) Data. NMR Biomed. 2016; 29(5):640–9
22. Wu D and Zhang J. Evidence of the Diffusion Time Dependence of Intravoxel Incoherent Motion in the Brain. Magn Reson Med. 2019; 82(6):2225–35
23. Gurney-Champion OJ, Rauh SS, Harrington K, Oelfke U, Laun FB and Wetscherek A. Optimal Acquisition Scheme for Flow-Compensated Intravoxel Incoherent Motion Diffusion-Weighted Imaging in the Abdomen: An Accurate and Precise Clinically Feasible Protocol. Magn Reson Med. 2020 Mar; 83(3):1003–15
24. Jiang L, Sun T, Liao Y, Sun Y, Qian Z, Zhang Y et al. Probing the Ballistic Microcirculation in Placenta Using Flow-Compensated and Non-Compensated Intravoxel Incoherent Motion Imaging. Magn Reson Med. 2021; 85(1):404–12

25. Liao Y, Sun T, Jiang L, Zhao Z, Liu T, Qian Z et al. Detecting Abnormal Placental Microvascular Flow in Maternal and Fetal Diseases Based on Flow-Compensated and Non-Compensated Intravoxel Incoherent Motion Imaging. Placenta. 2022 Mar 4; 119:17–23

26. He N, Li Z, Li X, Dai W, Peng C, Wu Y et al. Intravoxel Incoherent Motion Diffusion-Weighted Imaging Used to Detect Prostate Cancer and Stratify Tumor Grade: A Meta-Analysis. Front Oncol. 2020; 10:1623

27. Kooreman ES, van Houdt PJ, Keesman R, van Pelt VWJ, Nowee ME, Pos F et al. Daily Intravoxel Incoherent Motion (IVIM) In Prostate Cancer Patients During MR-Guided Radiotherapy-A Multicenter Study. Front Oncol. 2021; 11:705964

28. Ter Voert EEGW, Delso G, Porto M, Huellner M and Veit-Haibach P. Intravoxel Incoherent Motion Protocol Evaluation and Data Quality in Normal and Malignant Liver Tissue and Comparison to the Literature. Invest Radiol. 2016 Feb; 51(2):90

29. Jalnefjord O, Montelius M, Starck G and Ljungberg M. Optimization of B-Value Schemes for Estimation of the Diffusion Coefficient and the Perfusion Fraction with Segmented Intravoxel Incoherent Motion Model Fitting. Magn Reson Med. 2019; 82(4):1541–52

30. Merisaari H and Federau C. Signal to Noise and B-Value Analysis for Optimal Intra-Voxel Incoherent Motion Imaging in the Brain. PloS One. 2021; 16(9):e0257545

31. Zhang JL, Sigmund EE, Rusinek H, Chandarana H, Storey P, Chen Q et al. Optimization of B-Value Sampling for Diffusion-Weighted Imaging of the Kidney. Magn Reson Med. 2012; 67(1):89–97

32. Lemke A, Stieltjes B, Schad LR and Laun FB. Toward an Optimal Distribution of b Values for Intravoxel Incoherent Motion Imaging. Magn Reson Imaging. 2011 Jul 1; 29(6):766–76

33. Wurnig MC, Donati OF, Ulbrich E, Filli L, Kenkel D, Thoeny HC et al. Systematic Analysis of the Intravoxel Incoherent Motion Threshold Separating Perfusion and Diffusion Effects: Proposal of a Standardized Algorithm. Magn Reson Med. 2015; 74(5):1414–22

34. Fadnavis S, Endres S, Wen Q, Wu YC, Cheng H, Koudoro S et al. Bifurcated Topological Optimization for IVIM. Front Neurosci. 2021; 15:779025

35. DIPY : Docs 1.1.0. - Intravoxel Incoherent Motion. Available from: `https://dipy.org/` [Accessed on: 2023 Mar 27]

36. Pieper CC, Willinek WA, Meyer C, Ahmadzadehfar H, Kukuk GM, Sprinkart AM et al. Intravoxel Incoherent Motion Diffusion-Weighted MR Imaging for Prediction of Early Arterial Blood Flow Stasis in Radioembolization of Breast Cancer Liver Metastases. J Vasc Interv Radiol. 2016 Sep 1; 27(9):1320–8

37. Golub G and Pereyra V. Separable Nonlinear Least Squares: The Variable Projection Method and Its Applications. Inverse Probl. 2003 Feb; 19(2):R1

38. Farooq H, Xu J, Nam JW, Keefe DF, Yacoub E, Georgiou T et al. Microstructure Imaging of Crossing (MIX) White Matter Fibers from Diffusion MRI. Sci Rep. 2016 Dec 16; 6(1):38927

39. Fadnavis S, Reisert M, Farooq H, Afzali M, Cheng H, Amirbekian B et al. MicroLearn: Framework for Machine Learning, Reconstruction, Optimization and Microstructure Modeling. In: *Proceedings of the 27th Annual Meeting of ISMRM*. Montréal, Canada, 2019

40. Leporq B, Saint-Jalmes H, Rabrait C, Pilleul F, Guillaud O, Dumortier J et al. Optimization of Intra-Voxel Incoherent Motion Imaging at 3.0 Tesla for Fast Liver Examination. J Magn Reson Imaging. 2015; 41(5):1209–17

41. Karlsen OT, Verhagen R and Bovée WM. Parameter Estimation from Rician-distributed Data Sets Using a Maximum Likelihood Estimator: Application to T1 and Perfusion Measurements. Magn Reson Med. 1999; 41(3):614–23

42. Gudbjartsson H and Patz S. The Rician Distribution of Noisy MRI Data. Magn Reson Med. 1995 Dec; 34(6):910–4

43. McCann AJ, Workman A and McGrath C. A Quick and Robust Method for Measurement of Signal-to-Noise Ratio in MRI. Phys Med Biol. 2013 Jun 7; 58(11):3775–90

44. Determination of Signal-to-Noise Ratio (SNR) in Diagnostic Magnetic Resonance Imaging. NEMA MS 1-2008 (R2014). Rosslyn, VA: National Electrical Manufacturers Association, 2015 Apr 30

45. Coupé P, Manjón JV, Gedamu E, Arnold D, Robles M and Collins DL. Robust Rician Noise Estimation for MR Images. Med Image Anal. 2010 Aug 1; 14(4):483–93

46. Englund EK, Reiter DA, Shahidi B and Sigmund EE. Intravoxel Incoherent Motion Magnetic Resonance Imaging in Skeletal Muscle: Review and Future Directions. J Magn Reson Imaging. 2022; 55(4):988–1012

47. Paganelli C, Zampini MA, Morelli L, Buizza G, Fontana G, Anemoni L et al. Optimizing B-Values Schemes for Diffusion MRI of the Brain with Segmented Intravoxel Incoherent Motion (IVIM) Model. J Appl Clin Med Phys. 2023 :e13986

48. Rydhög A, Pasternak O, Ståhlberg F, Ahlgren A, Knutsson L and Wirestam R. Estimation of Diffusion, Perfusion and Fractional Volumes Using a Multi-Compartment Relaxation-Compensated Intravoxel Incoherent Motion (IVIM) Signal Model. Eur J Radiol Open. 2019 May 24; 6:198–205

49. Federau C and O'Brien K. Increased Brain Perfusion Contrast with T2-prepared Intravoxel Incoherent Motion (T2prep IVIM) MRI. NMR Biomed. 2015 Jan; 28(1):9–16

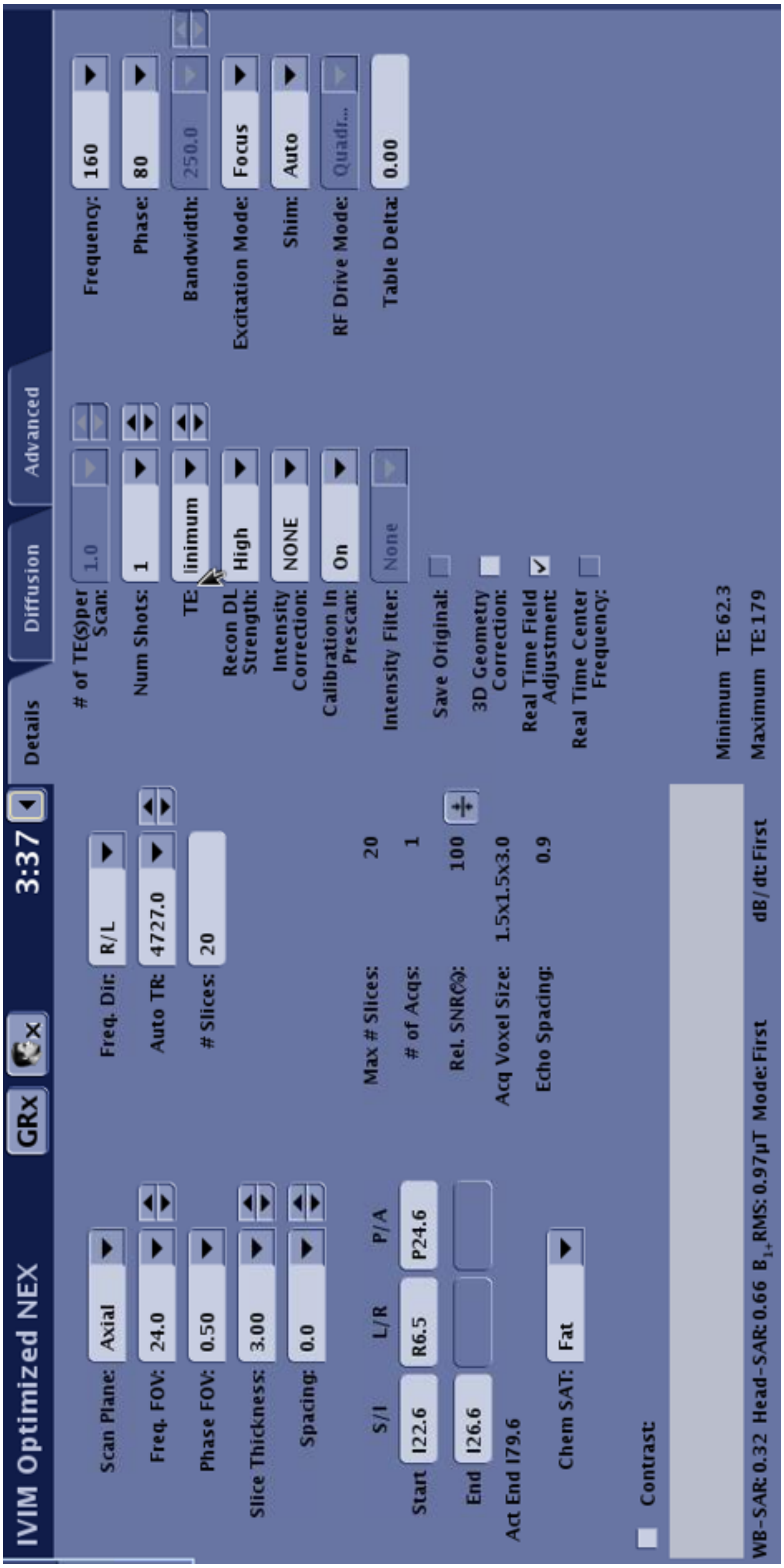
50. Chatterjee A, Bourne RM, Wang S, Devaraj A, Gallan AJ, Antic T et al. Diagnosis of Prostate Cancer with Noninvasive Estimation of Prostate Tissue Composition by Using Hybrid Multidimensional MR Imaging: A Feasibility Study. Radiol. 2018 Jun; 287(3):864–73

51. Chatterjee A, Mercado C, Bourne RM, Yousuf A, Hess B, Antic T et al. Validation of Prostate Tissue Composition by Using Hybrid Multidimensional MRI: Correlation with Histologic Findings. Radiol. 2022 Feb; 302(2):368–77

52. OSIPI TF2.4: IVIM MRI Code Collection. Available from: `https://github.com/OSIPI/TF2.4_IVIM-MRI_CodeCollection` [Accessed on: 2023 Jun 22]

A Appendix: Pulse Sequence Settings

Below are screenshots with the settings of the used pulse sequences on the GE 3T system.



IVIM Optimized NEX

3:37

Scan Plane: Axial
Freq. FOV: 24.0
Phase FOV: 0.50
Slice Thickness: 3.00
Spacing: 0.0

	S/I	L/R	P/A
Start	I22.6	R6.5	P24.6
End	I26.6		

Act End I79.6

Chem SAT: Fat

☐ Contrast:

Freq. Dir: R/L
Auto TR: 4727.0
Slices: 20

Max # Slices: 20
# of Acqs: 1
Rel. SNR(%): 100
Acq Voxel Size: 1.5x1.5x3.0
Echo Spacing: 0.9

# of TE(s)per Scan: 1.0
Num Shots: 1
TE: linimum
Recon DL Strength: High
Intensity Correction: NONE
Calibration In Prescan: On
Intensity Filter: None
Save Original: ☐
3D Geometry Correction: ☐
Real Time Field Adjustment: ☑
Real Time Center Frequency: ☐

Frequency: 160
Phase: 80
Bandwidth: 250.0
Excitation Mode: Focus
Shim: Auto
RF Drive Mode: Quadr...
Table Delta: 0.00

Minimum TE: 62.3
Maximum TE: 179

WB-SAR: 0.32 Head-SAR: 0.66 B_{1+}RMS: 0.97µT Mode: First dB/dt: First

IVIM Optimized NEX

Scan Plane: Axial
Freq. FOV: 24.0
Phase FOV: 0.50
Slice Thickness: 3.00
Spacing: 0.0

	S/I	L/R	P/A
Start	I22.6	R6.5	P24.6
End	I26.6		

Act End I79.6

Chem SAT: Fat

☐ Contrast:

Freq. Dir: R/L
Auto TR: 4727.0
Slices: 20

Max # Slices: 20
# of Acqs: 1
Rel. SNR(%): 100
Acq Voxel Size: 1.5x1.5x3.0
Echo Spacing: 0.9

b-values: 3
NEX for T2: 6.00
Diffusion Direction: ALL
# of Diffusion Directions: 3
# of T2 Images: 1
Synthetic b-values: 0

#	b-value	NEX:
1	50.0	3.0
2	240.0	5.0
3	800.0	5.0

Recon All Images: ☐
Optimize TE: ☑
Dual Spin Echo: ☐

WB-SAR: 0.32 Head-SAR: 0.66 B_{1+}RMS: 0.97µT Mode: First dB/dt: First

Scan Plane: Axial
Freq. FOV: 24.0
Phase FOV: 0.50
Slice Thickness: 3.00
Spacing: 0.0

	S/I	L/R	P/A
Start	I22.6	R6.5	P24.6
End	I26.6		

Act End I79.6

Chem SAT: Fat

☐ Contrast:

Freq. Dir: R/L
Auto TR: 4727.0
Slices: 20

Max # Slices: 20
# of Acqs: 1
Rel. SNR(%): 100
Acq Voxel Size: 1.5x1.5x3.0
Echo Spacing: 0.9

WB-SAR: 0.32 Head-SAR: 0.66 B_{1+}RMS: 0.97µT Mode: First dB/dt: First

User Control Variables

			Min	Max
CV0	Ramp Sampling (1=on, 0=off):	1.00	0.0	1.0
CV5	Recon Type (0=Zero Filling, 1=Homodyne):	1.00	0.0	1.0
CV7	Gradient Optimization for Diffusion ALL (1=on, 0=off):	1.00	0.0	1.0
CV9	Shim Volume Mode (0=Default, 1=Breast):	0.00	0.0	2.0

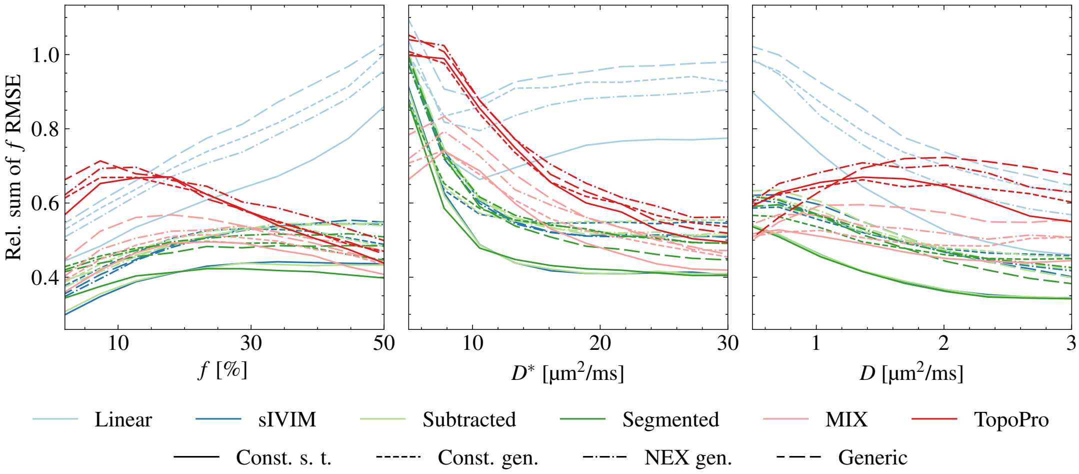
TR Range for AutoTR
Minimum TR: 2000.0
Maximum TR: 17000.0

B Appendix: RMSE Trends

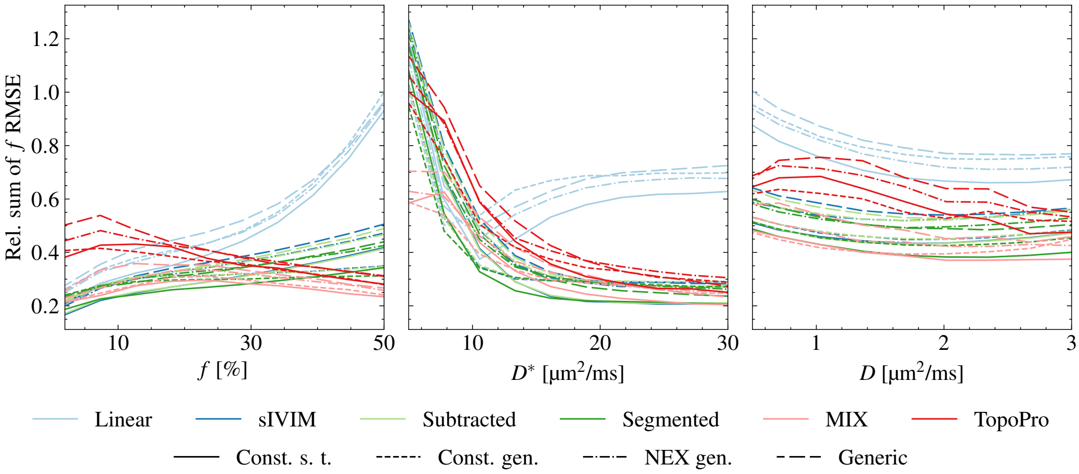
The RMSE maps presented in Appendix C were summed along different axes. The resulting curves for each fit and b-value scheme is presented below.

B.1 RMSE Trends for f

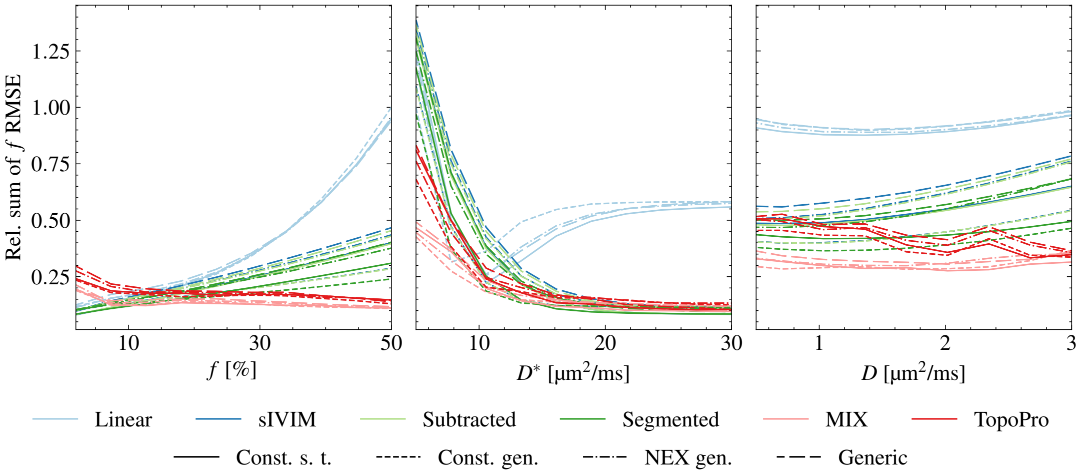
RMSE trends of f at SNR = 3



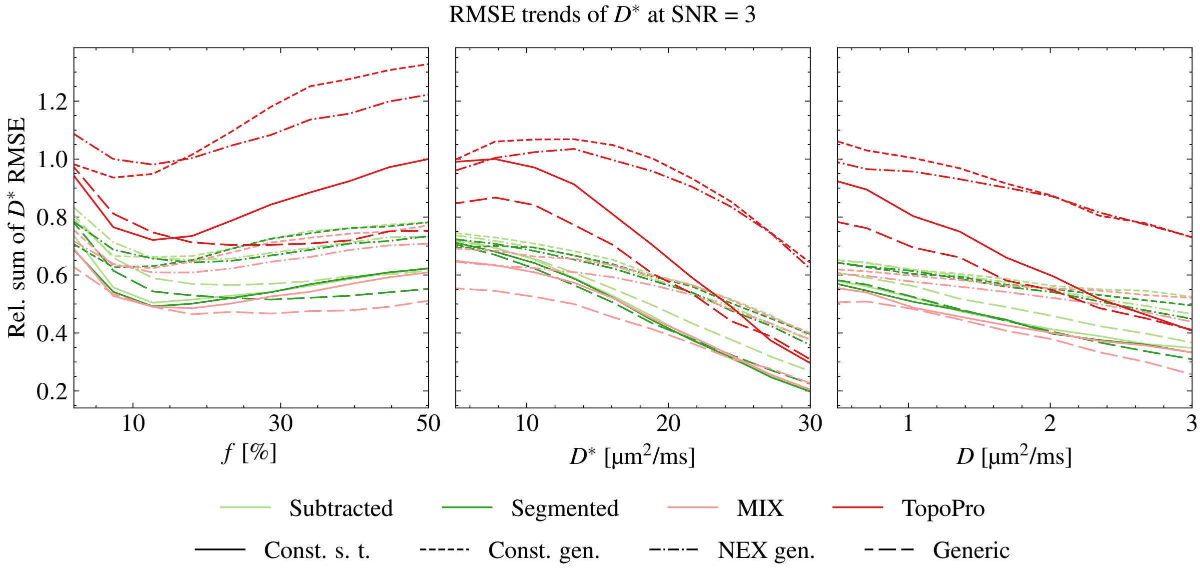
RMSE trends of f at SNR = 10

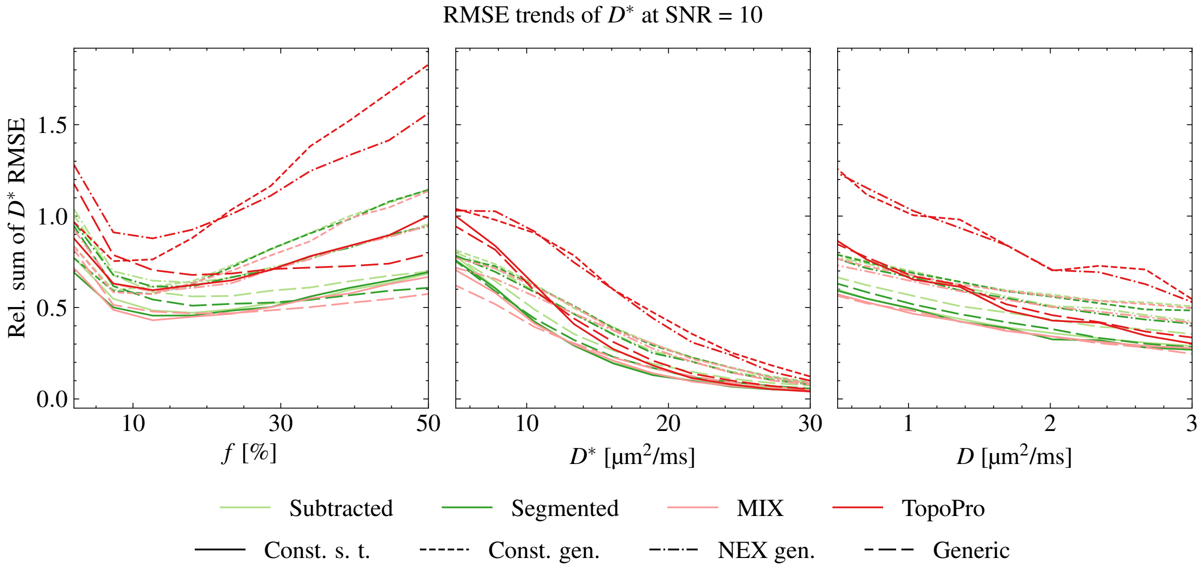


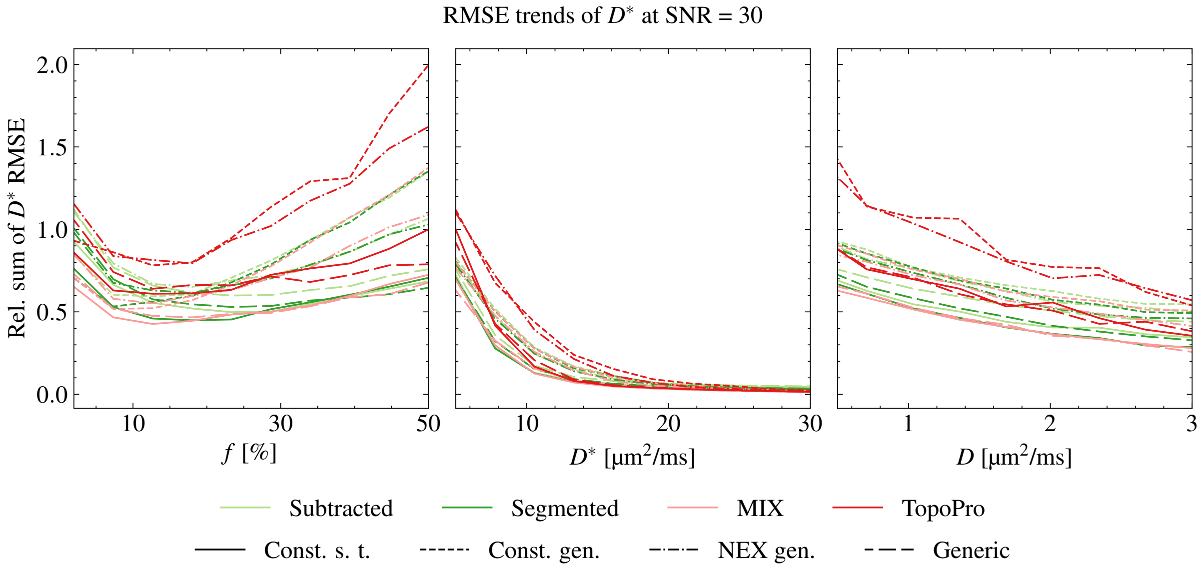
RMSE trends of f at SNR = 30



B.2 RMSE Trends for D^*

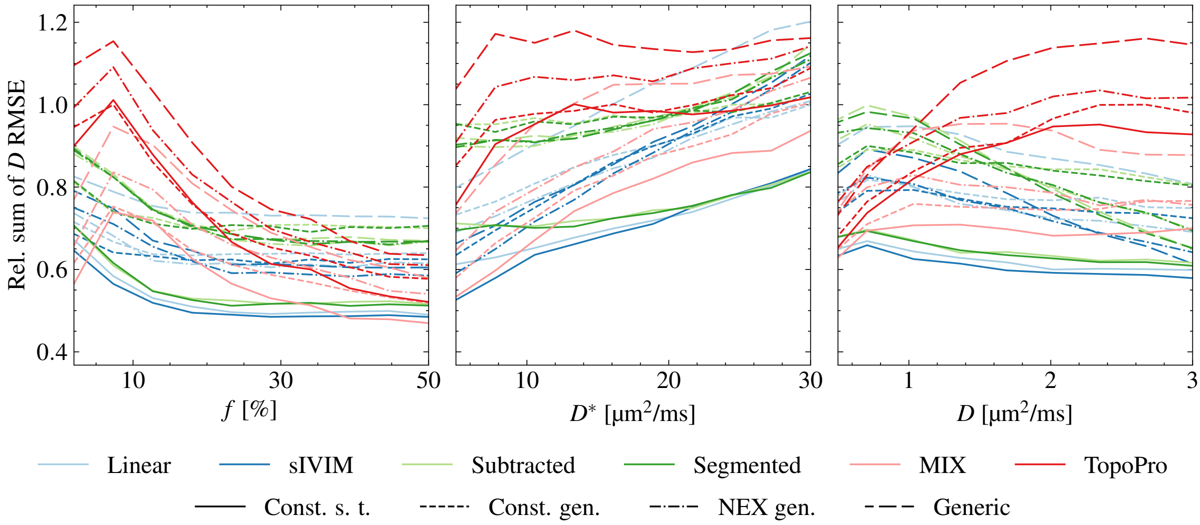

RMSE trends of D^* at SNR = 3


RMSE trends of D^* at SNR = 10

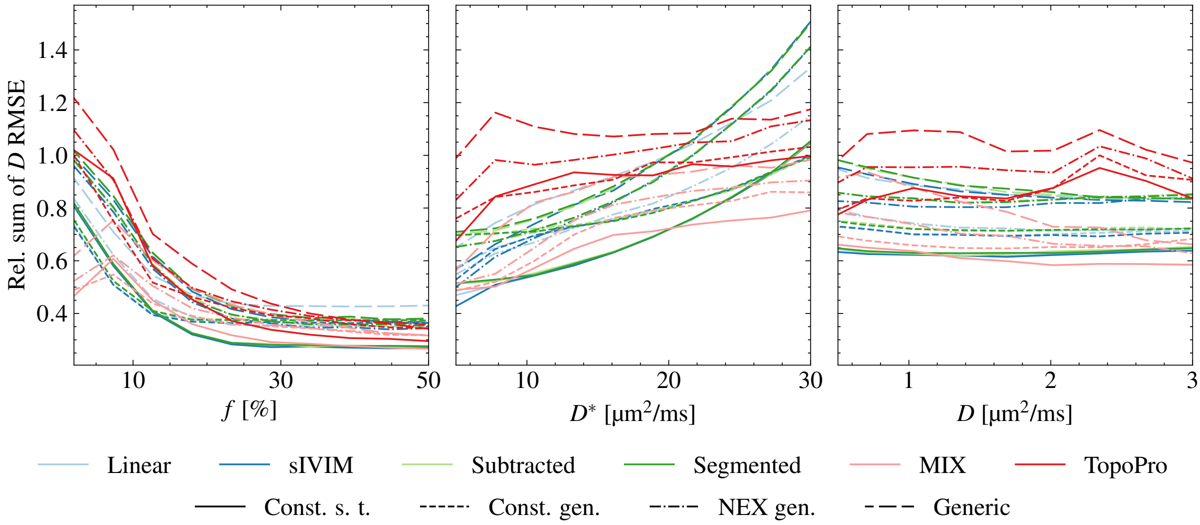

RMSE trends of D^* at SNR = 30

B.3 RMSE Trends for D

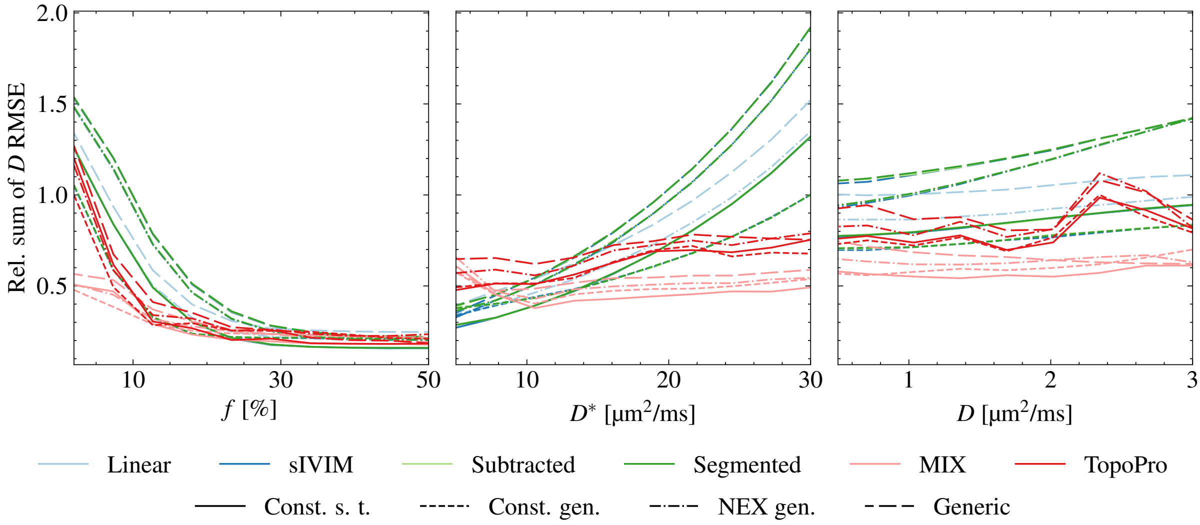
RMSE trends of D at SNR = 3



RMSE trends of D at SNR = 10



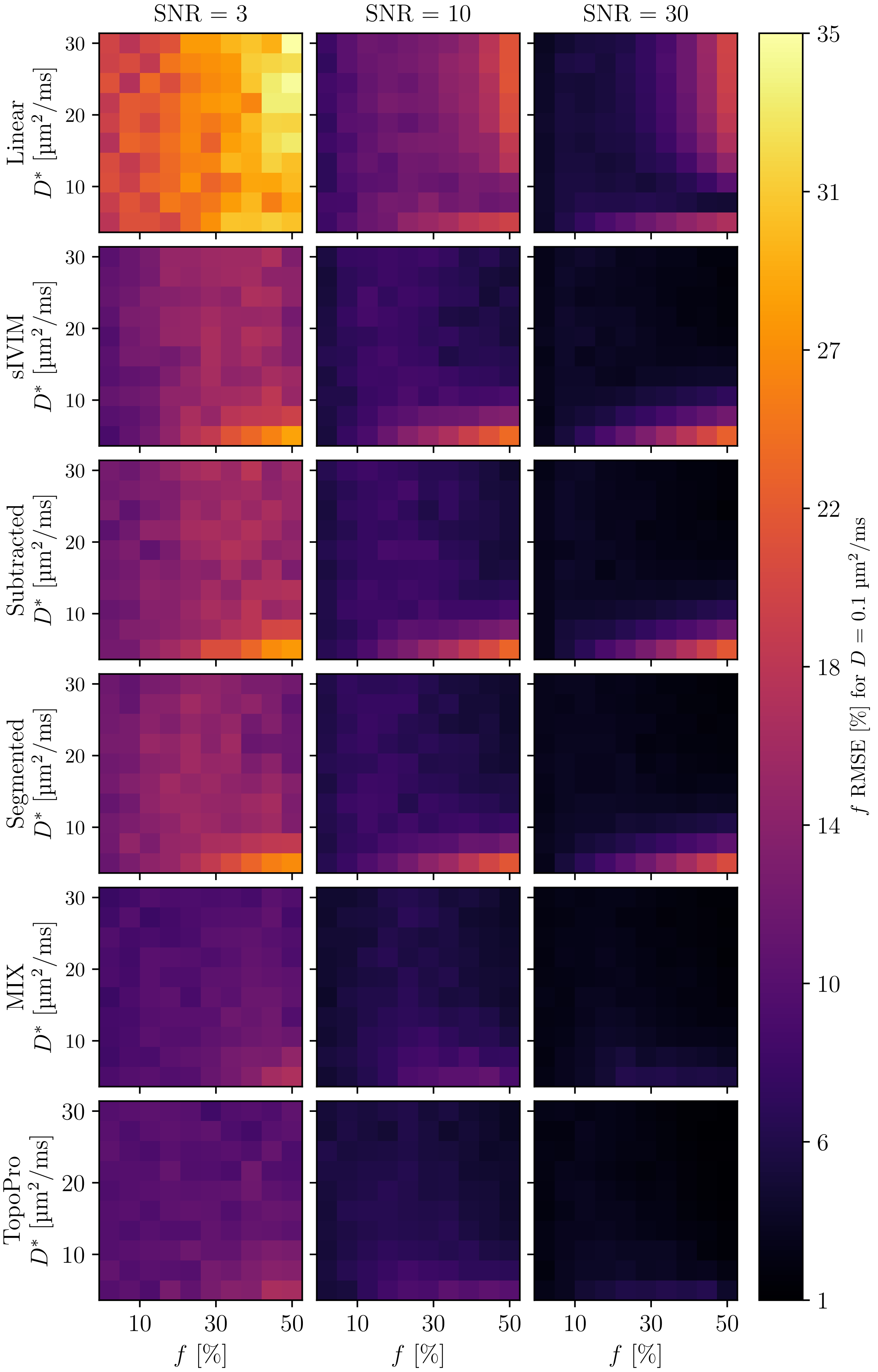
RMSE trends of D at SNR = 30

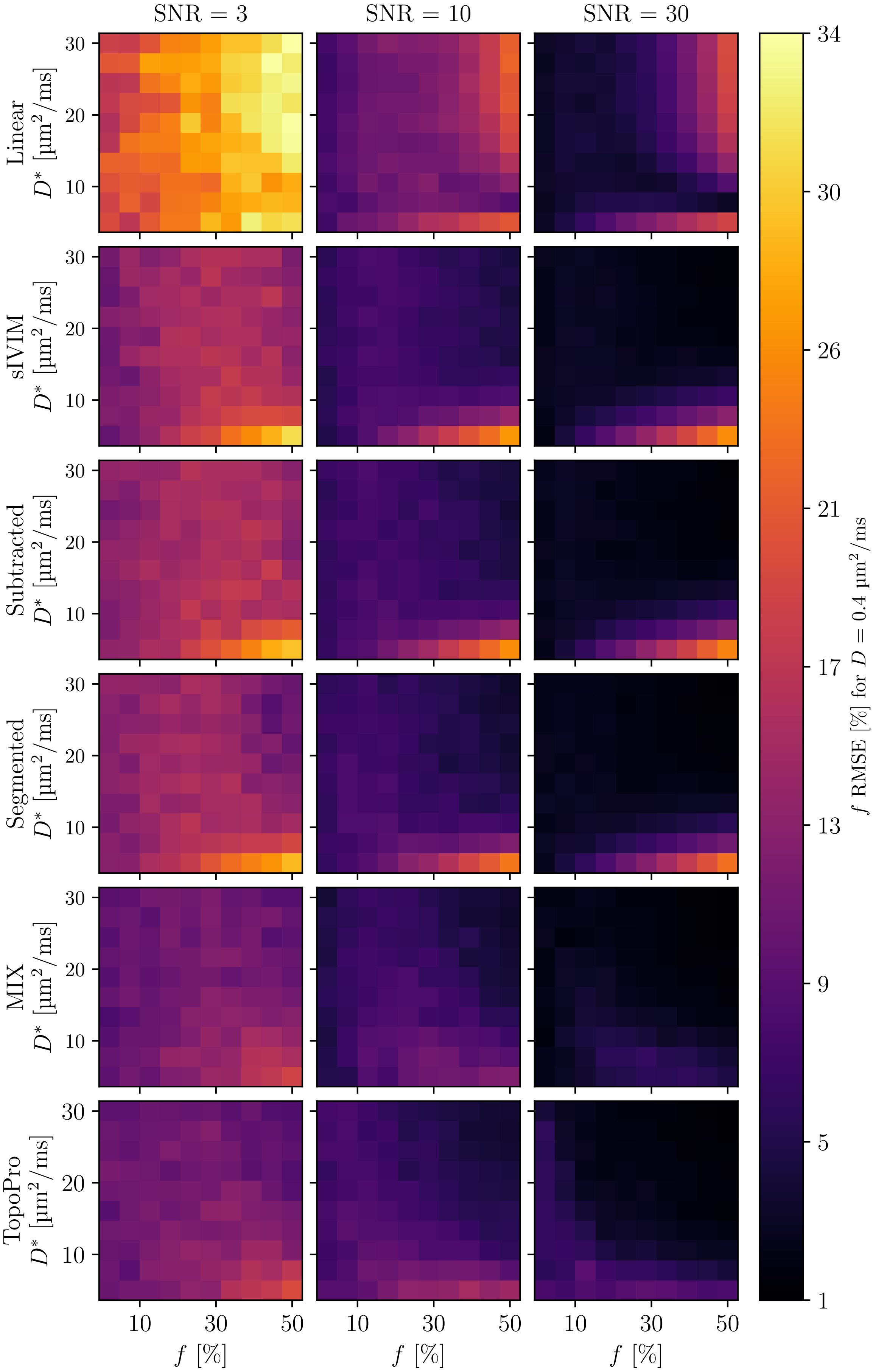


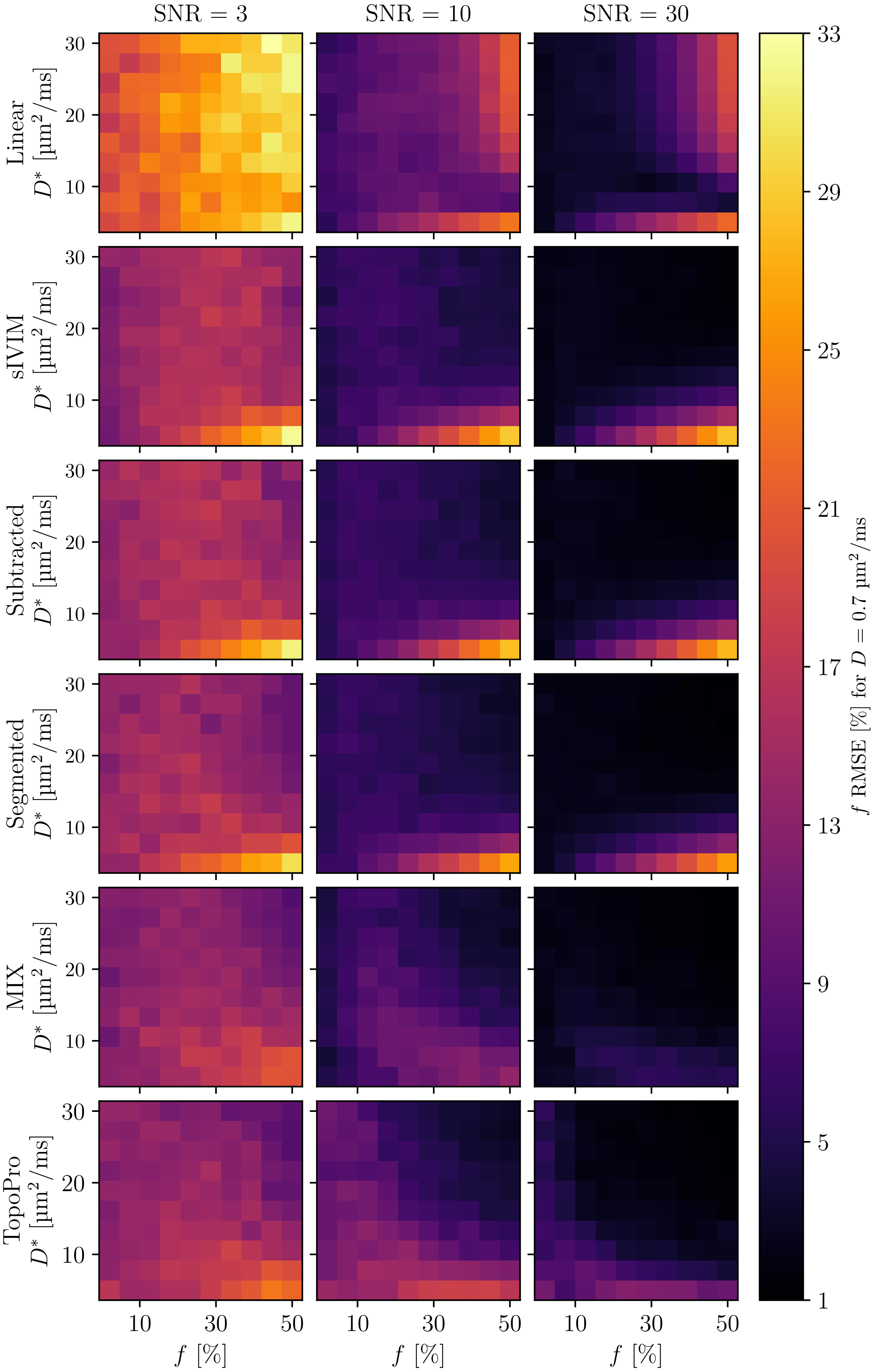
C Appendix: RMSE Maps

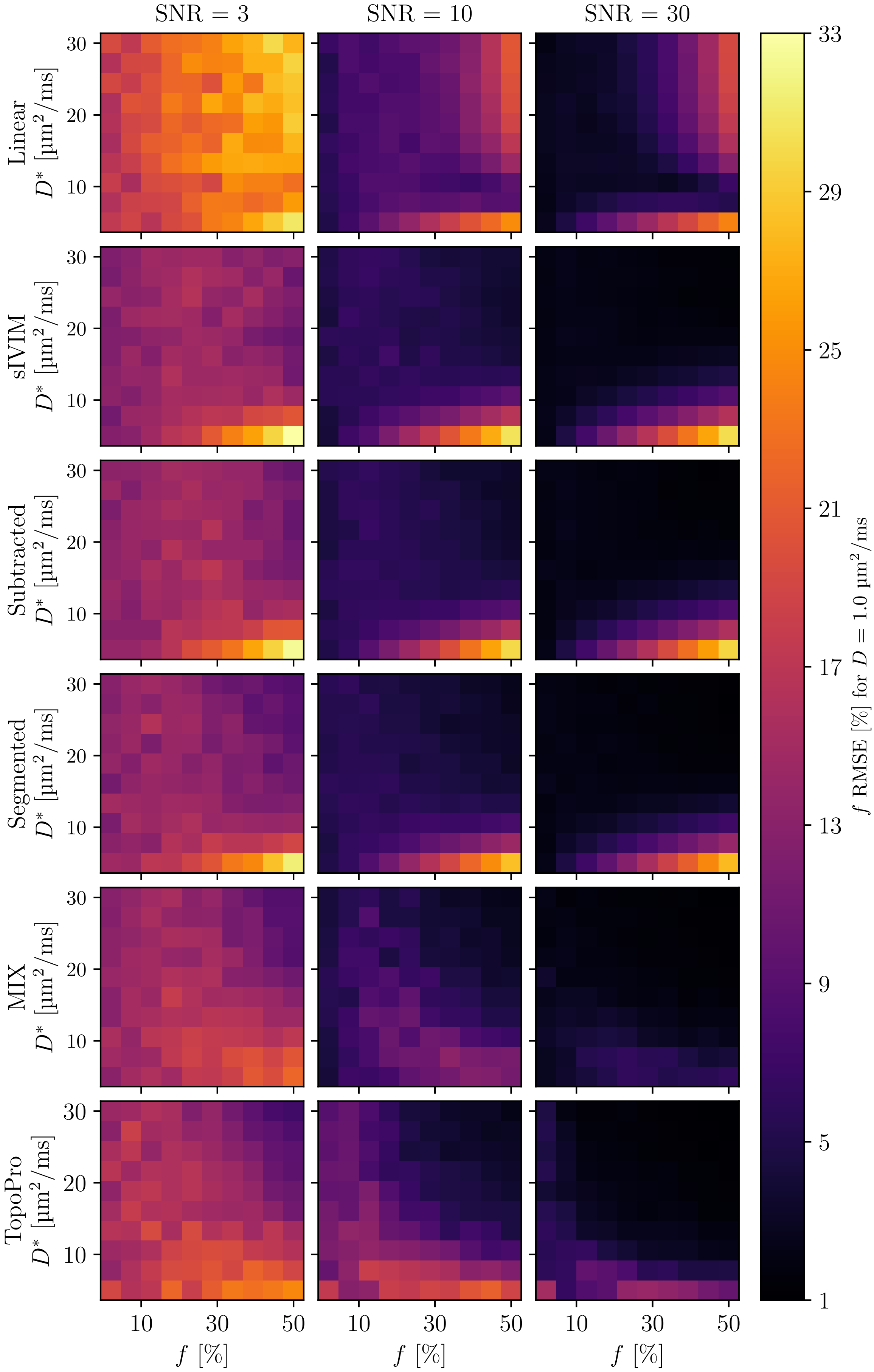
C.1 RMSE of f

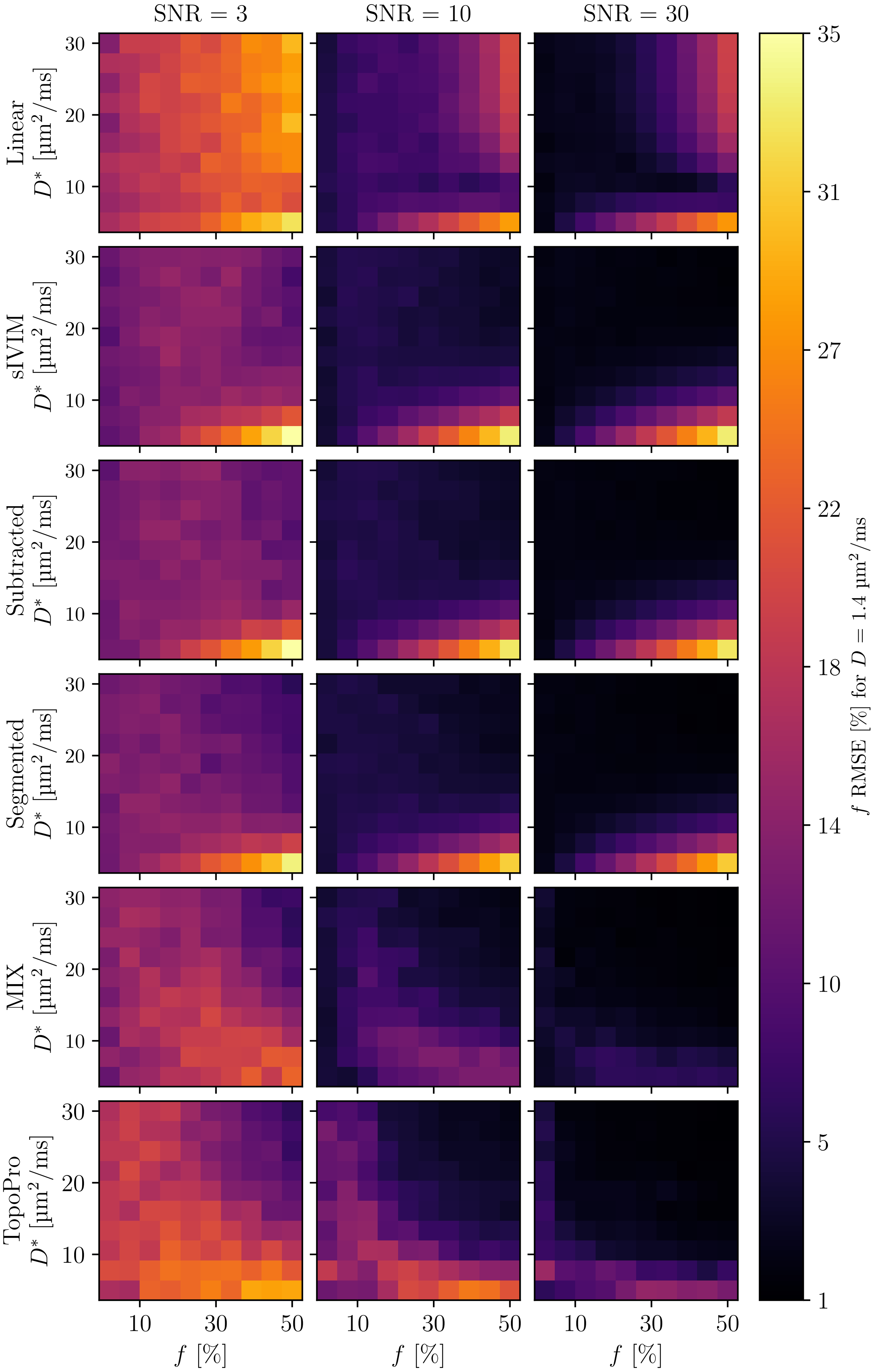
In this section, the RMSE maps of the f estimates for varying ground truth values of D in the slice direction are presented. The RMSE is presented in absolute units, i.e. the percentage unit of the maps should not be interpreted as a relative error.

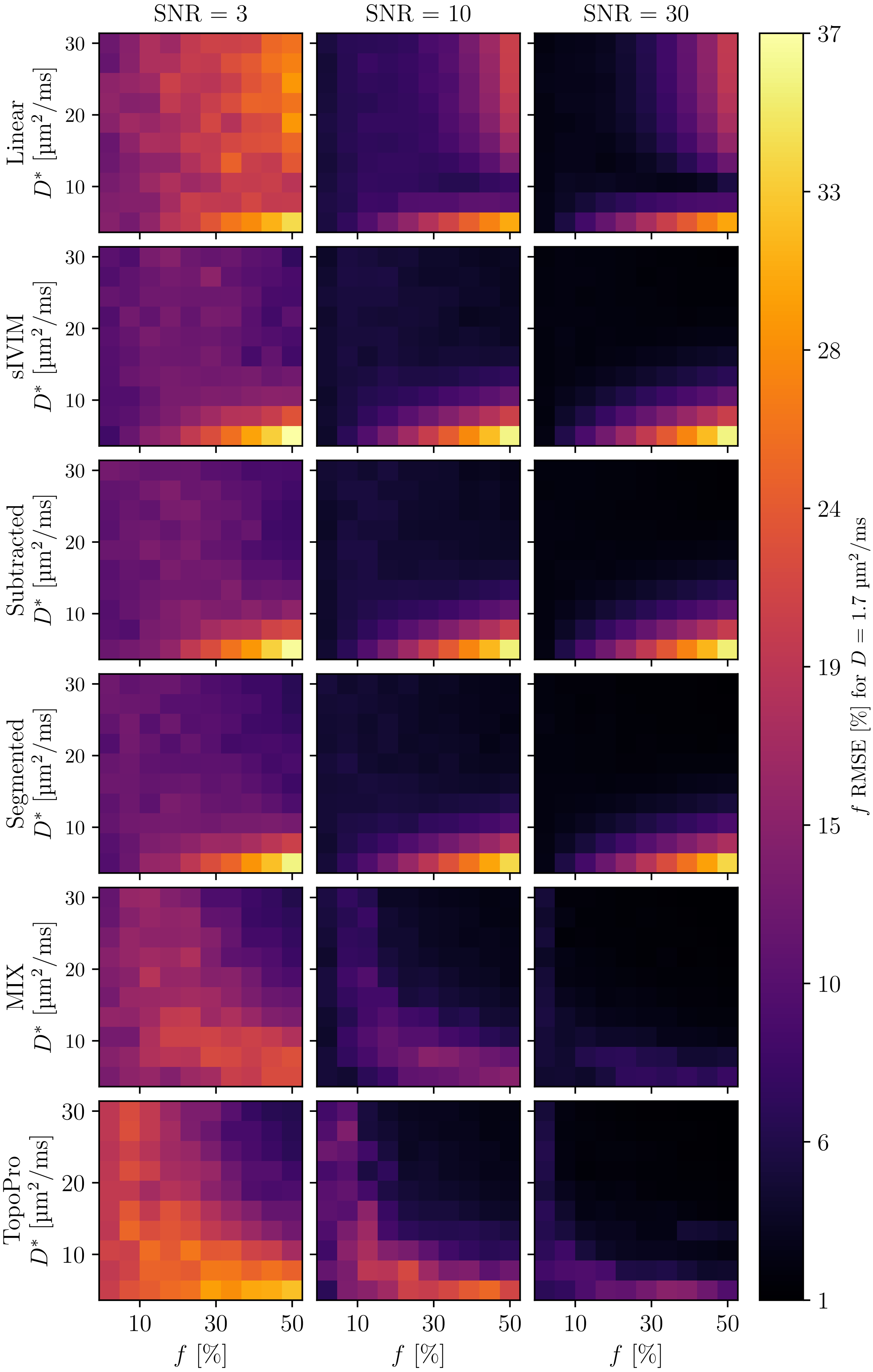

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
10
20
30
f [%]
10
30
50
f RMSE [%] for $D = 0.1$ µm^2/ms
1
6
10
14
18
22
27
31
35

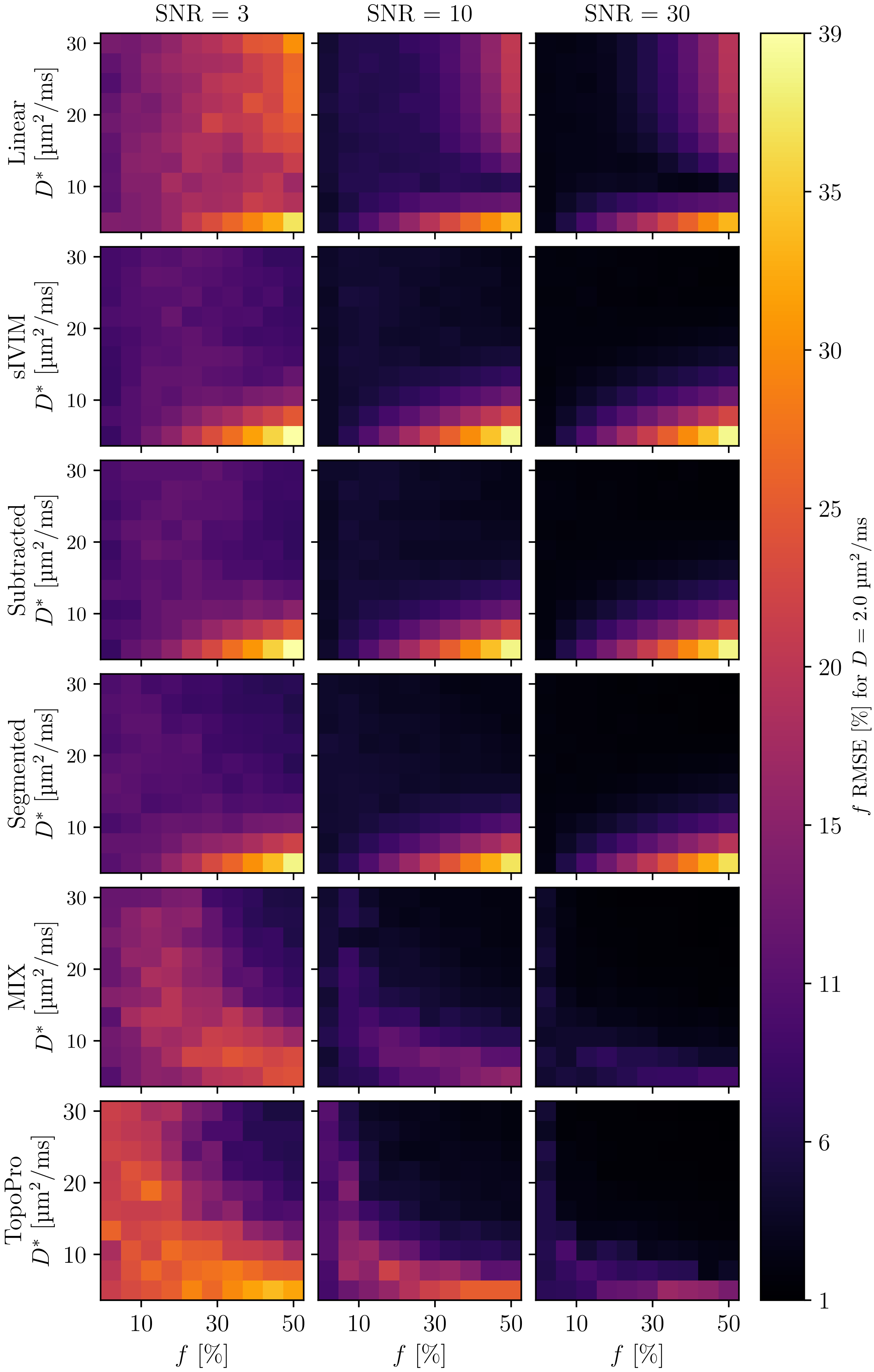

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
10
20
30
10
30
50
f [%]
f RMSE [%] for $D = 0.4$ μm^2/ms
1
5
9
13
17
21
26
30
34

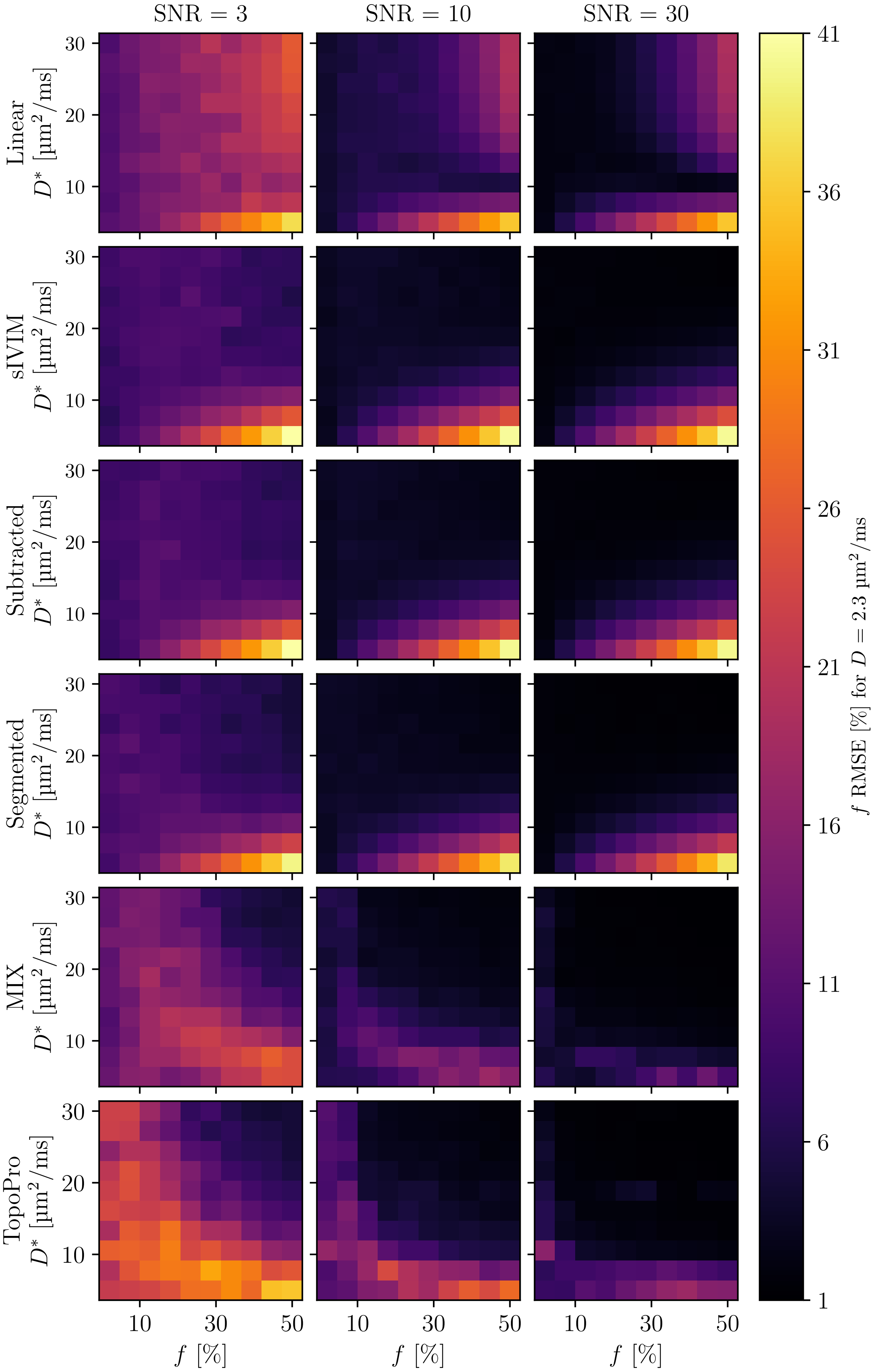

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
10
30
50
f [%]
f RMSE [%] for $D = 0.7$ µm^2/ms
33
29
25
21
17
13
9
5
1

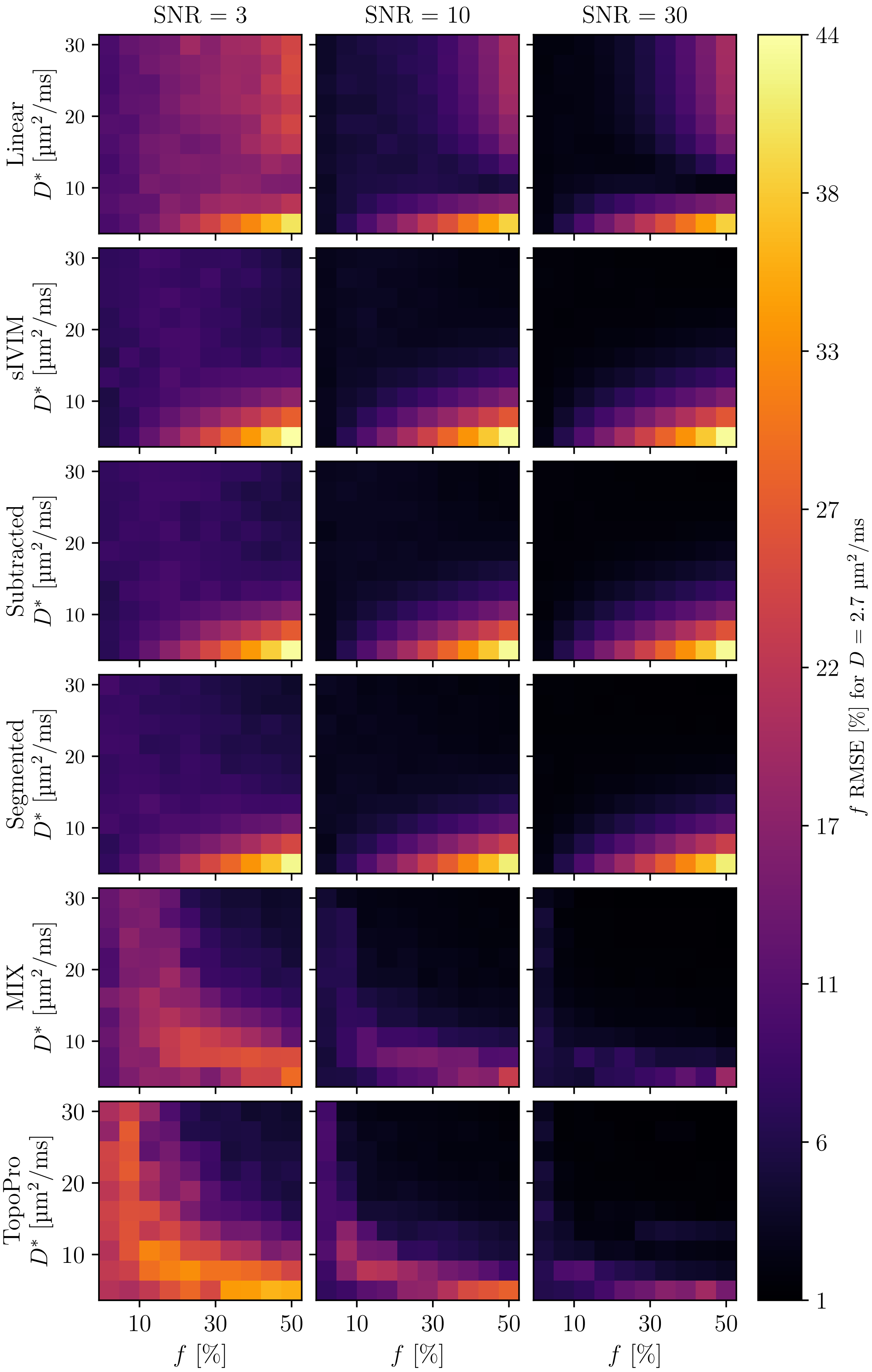

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
10
20
30
10
30
50
f [%]
1
5
9
13
17
21
25
29
33
f RMSE [%] for $D = 1.0$ µm^2/ms

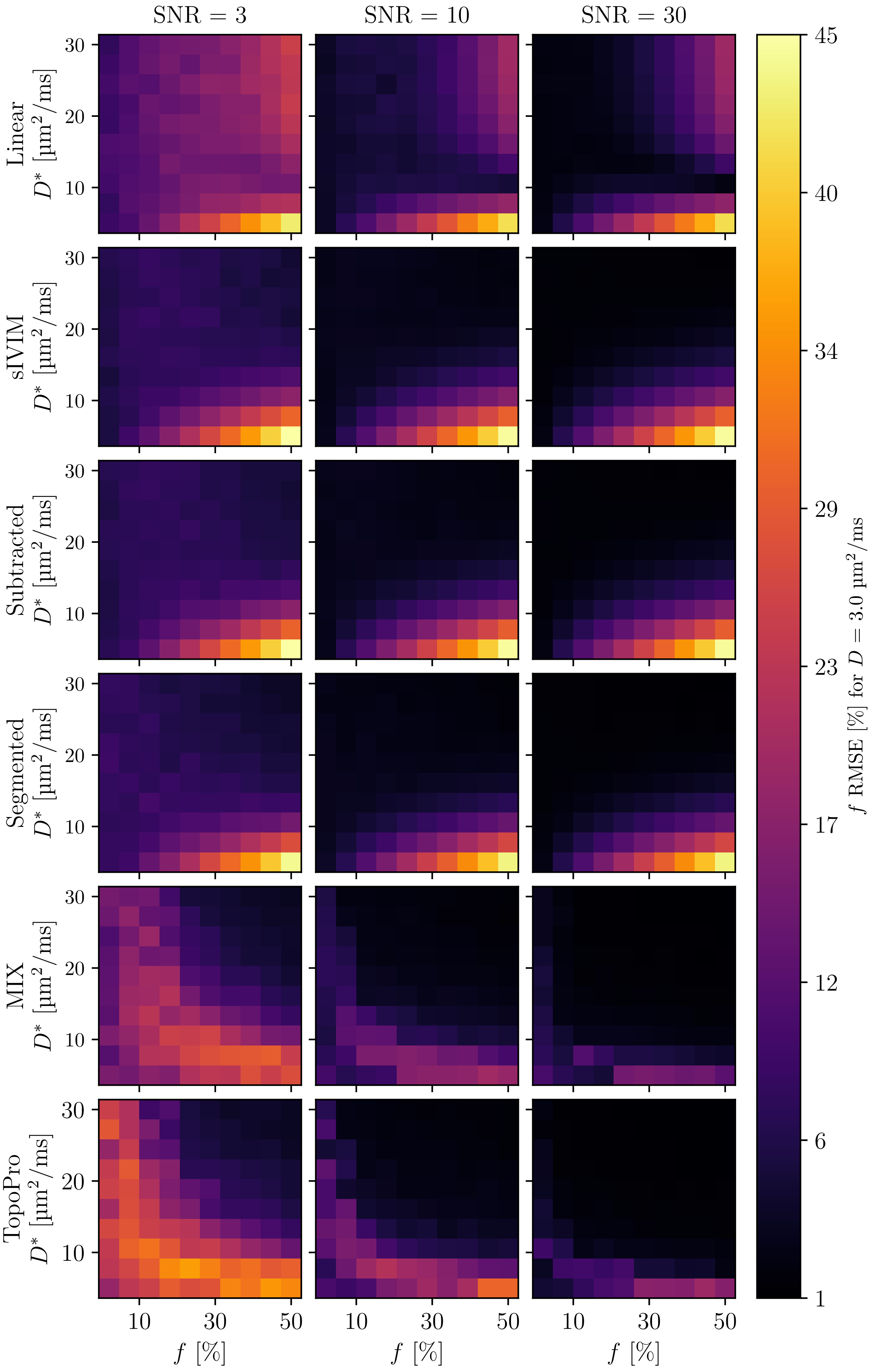

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
30
20
10
10
30
50
f [%]
35
31
27
22
18
14
10
5
1
f RMSE [%] for D = 1.4 μm^2/ms


SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
10
30
50
f [%]
f RMSE [%] for D = 1.7 µm^2/ms
37
33
28
24
19
15
10
6
1


SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
10
30
50
f [%]
f RMSE [%] for $D = 2.0$ µm^2/ms
39
35
30
25
20
15
11
6
1

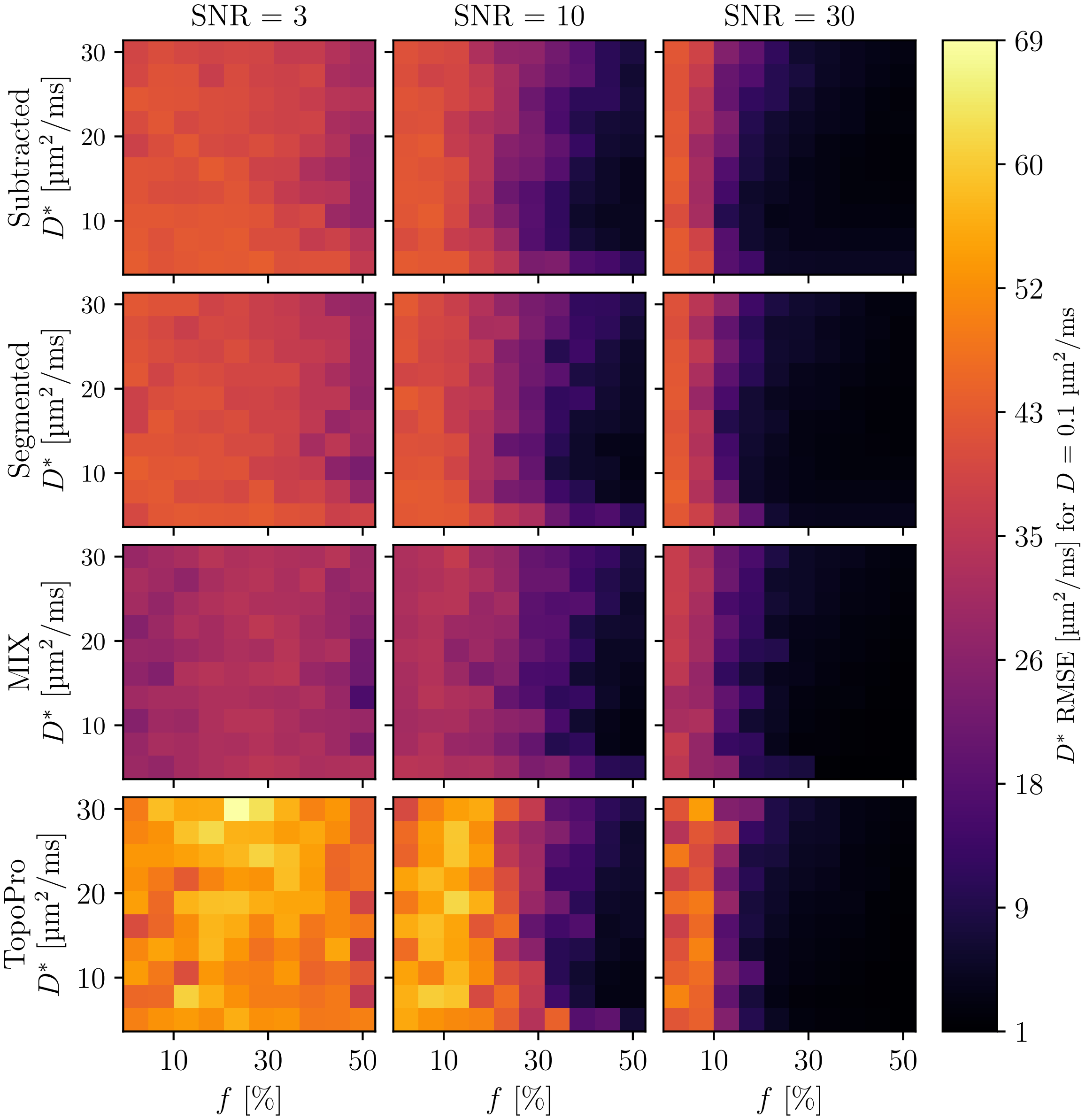

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
10
20
30
10
30
50
f [%]
f RMSE [%] for $D = 2.3$ μm^2/ms
1
6
11
16
21
26
31
36
41

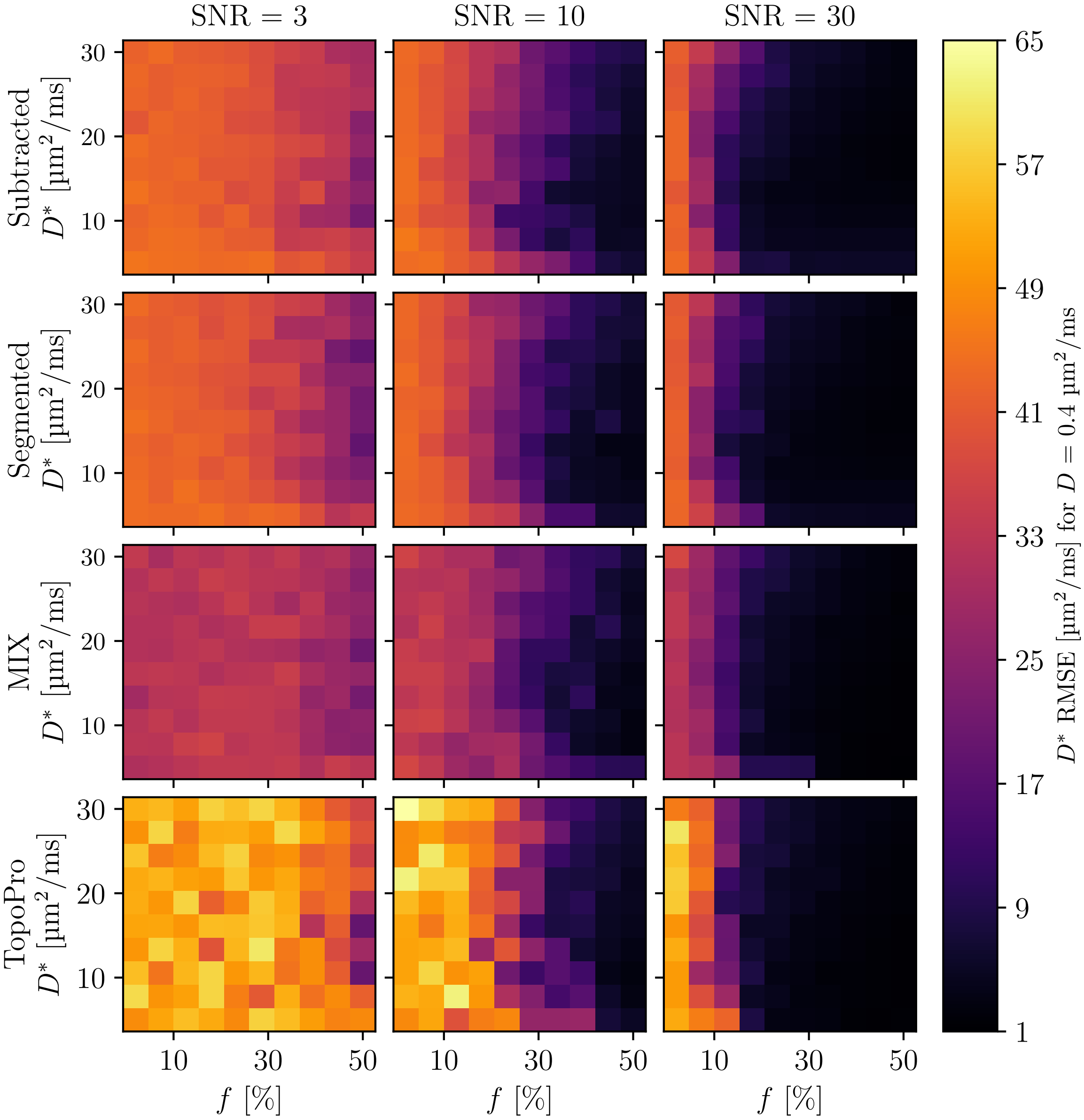

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
10
20
30
f [%]
10
30
50
44
38
33
27
22
17
11
6
1
f RMSE [%] for $D = 2.7$ µm^2/ms

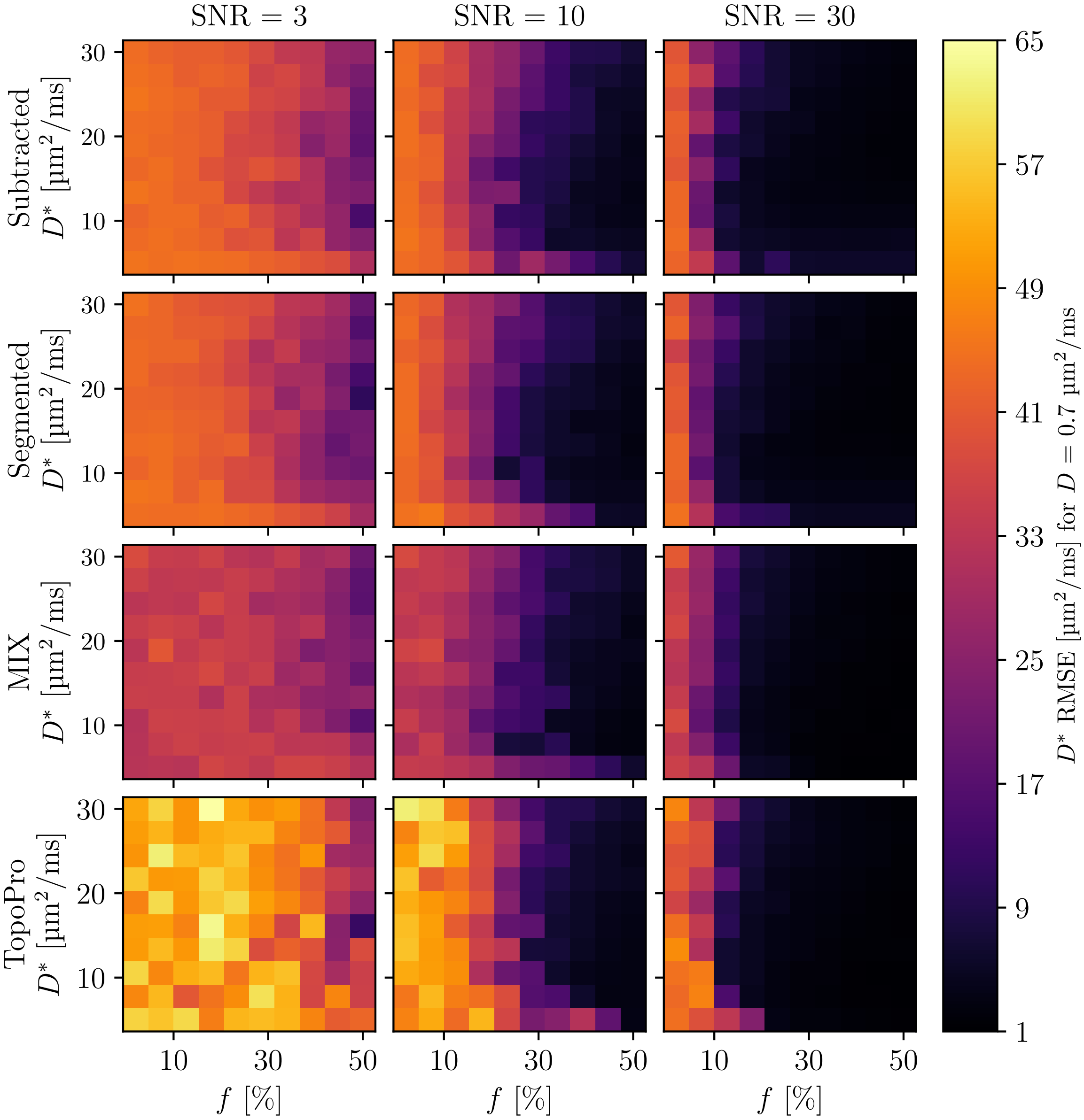

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
30
20
10
10
30
50
f [%]
45
40
34
29
23
17
12
6
1
f RMSE [%] for D = 3.0 μm^2/ms

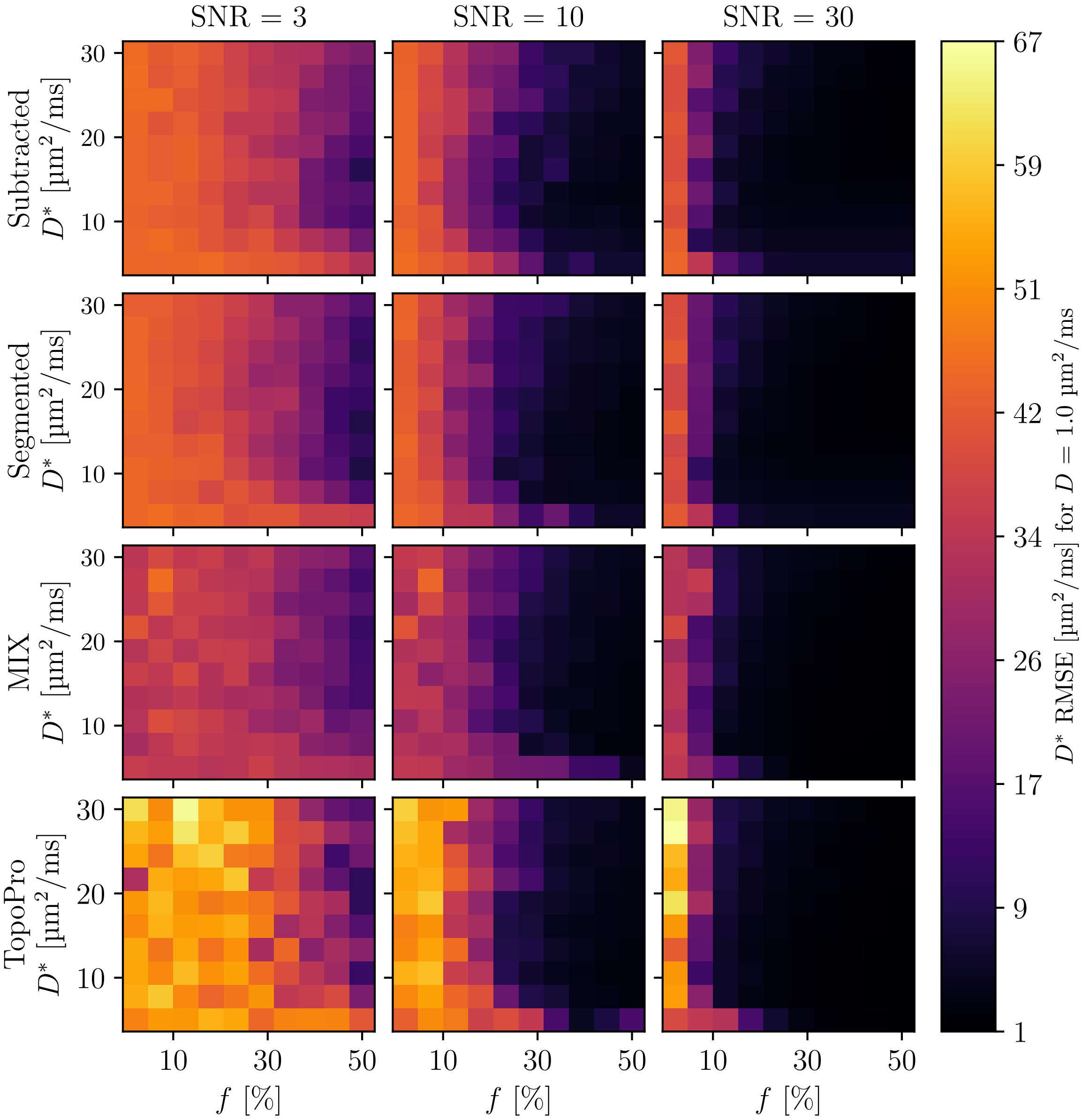
C.2 RMSE of D^*

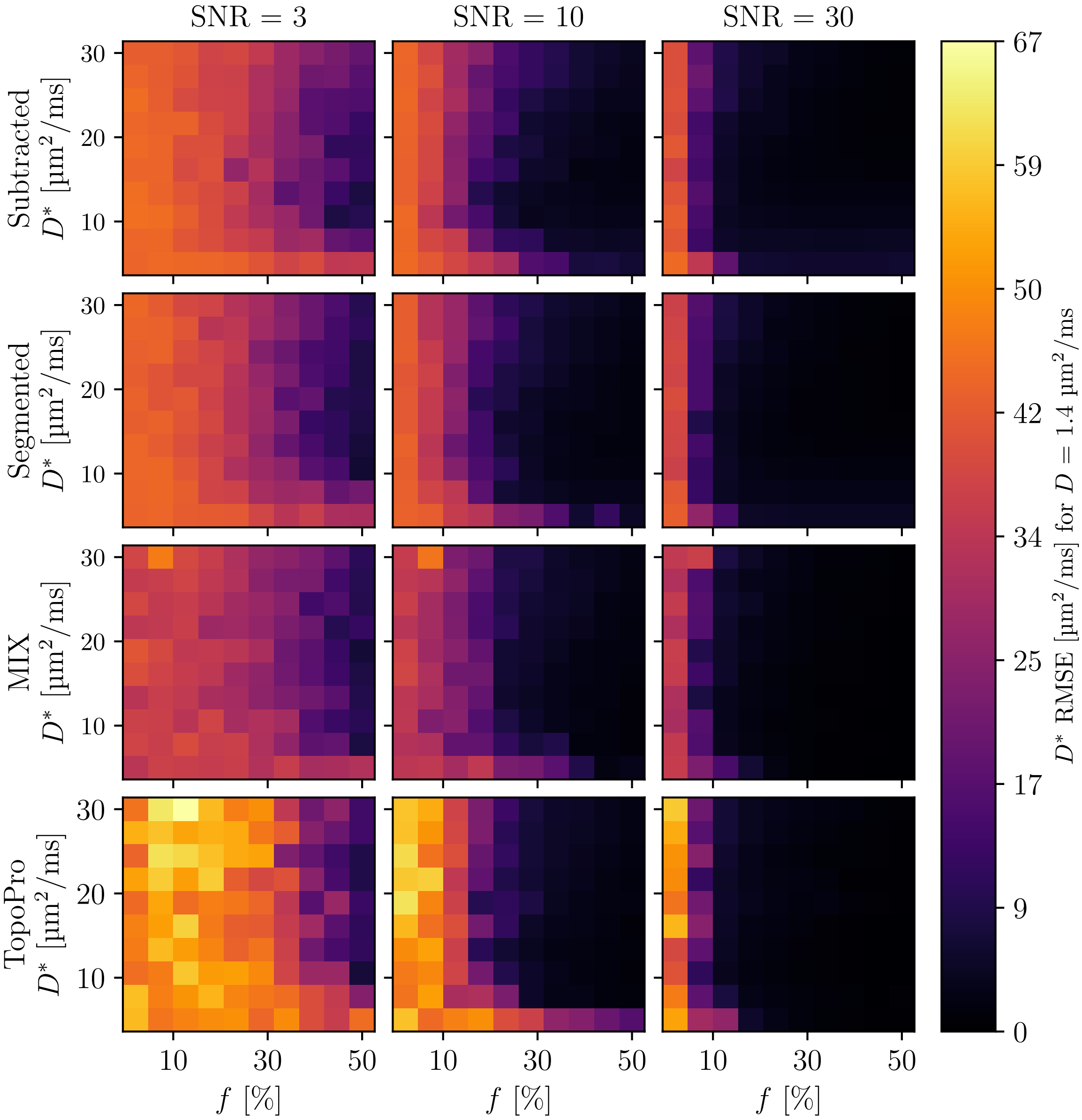
In this section, the RMSE maps of the D^* estimates for varying ground truth values of D in the slice direction are presented.

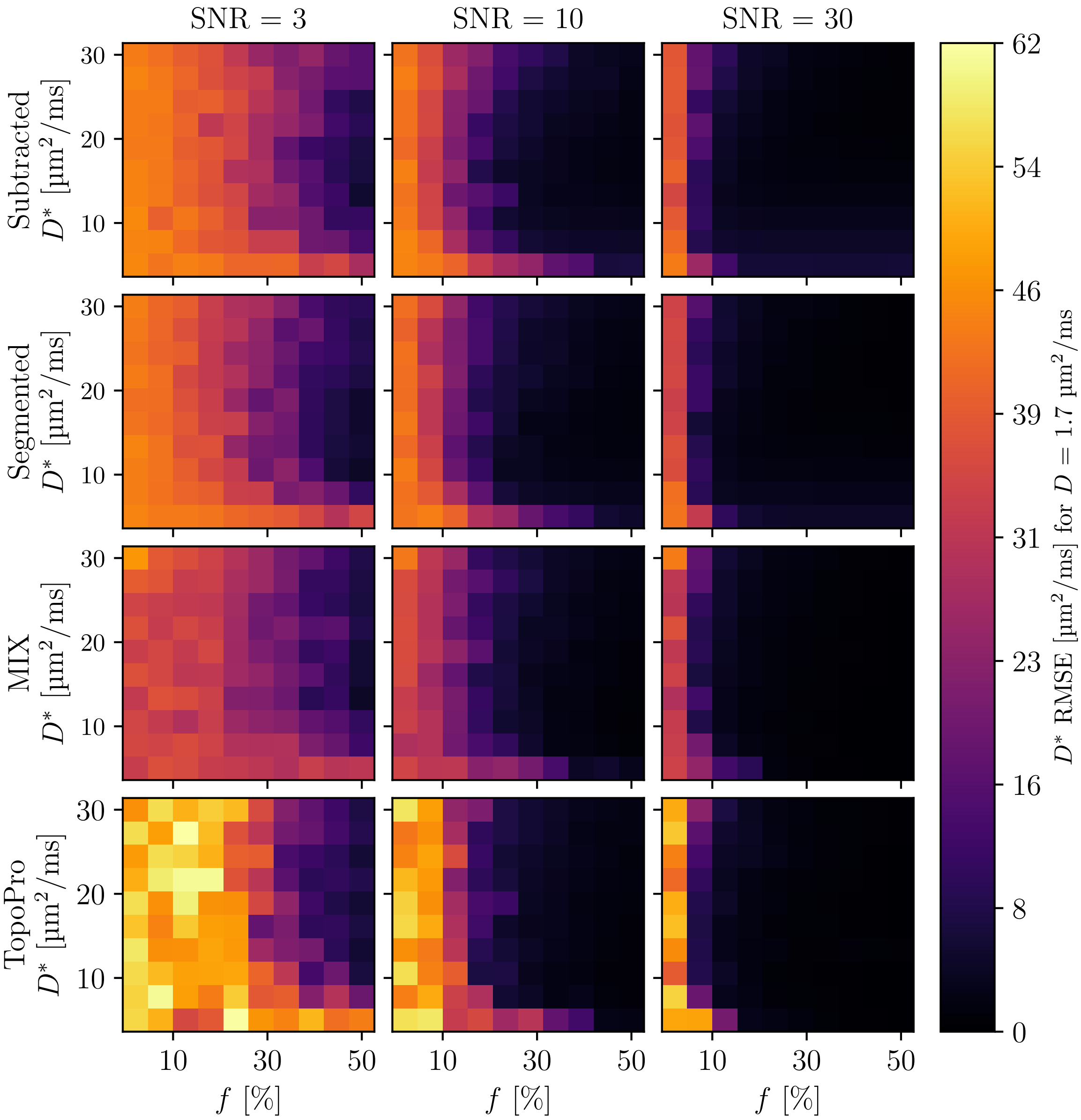

SNR = 3
SNR = 10
SNR = 30
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
30
20
10
10
30
50
f [%]
D^* RMSE [μm^2/ms] for $D = 0.1$ μm^2/ms
69
60
52
43
35
26
18
9
1

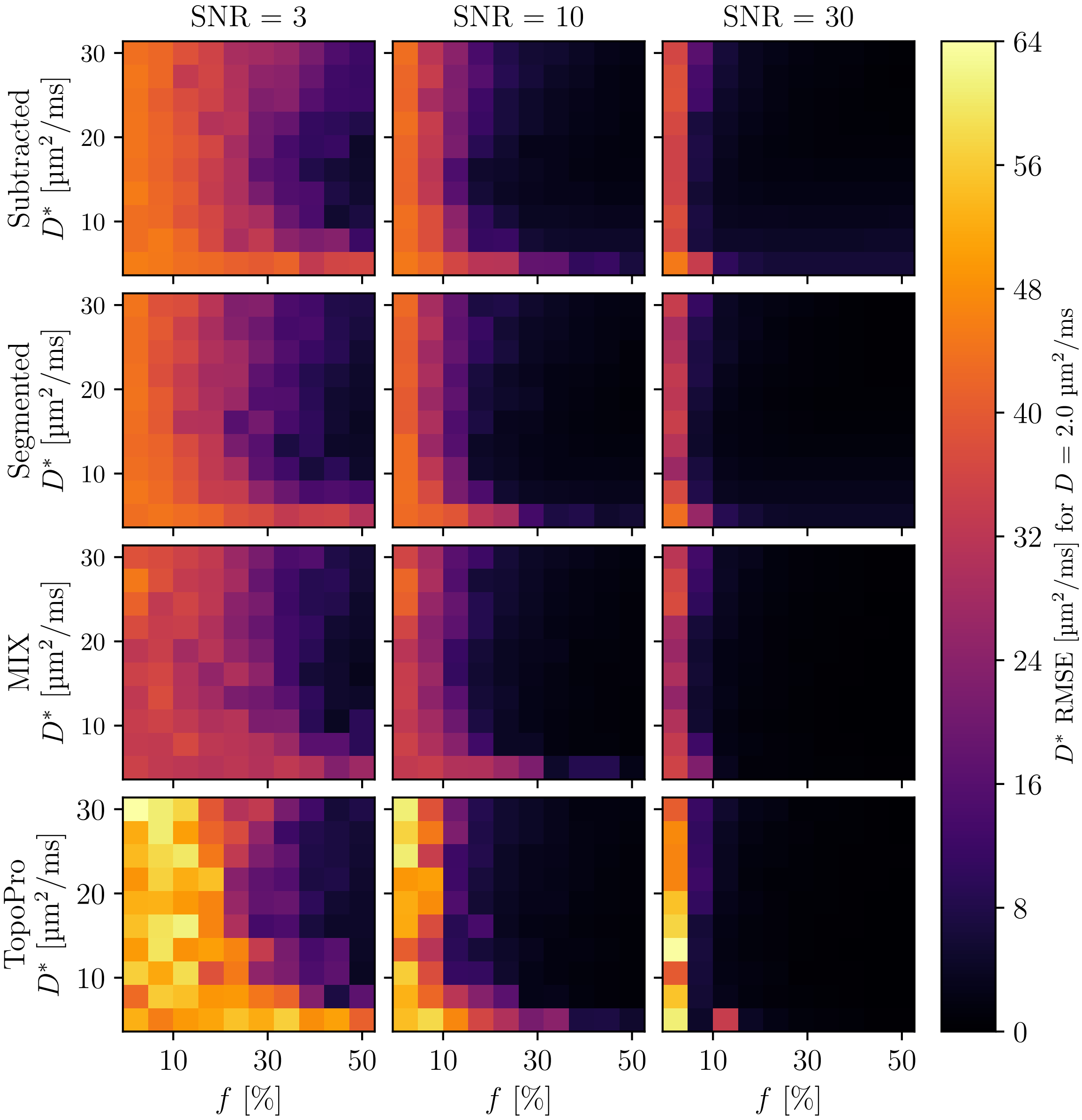

SNR = 3
SNR = 10
SNR = 30
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
10
20
30
f [%]
10
30
50
D^* RMSE [μm^2/ms] for $D = 0.4$ μm^2/ms
65
57
49
41
33
25
17
9
1

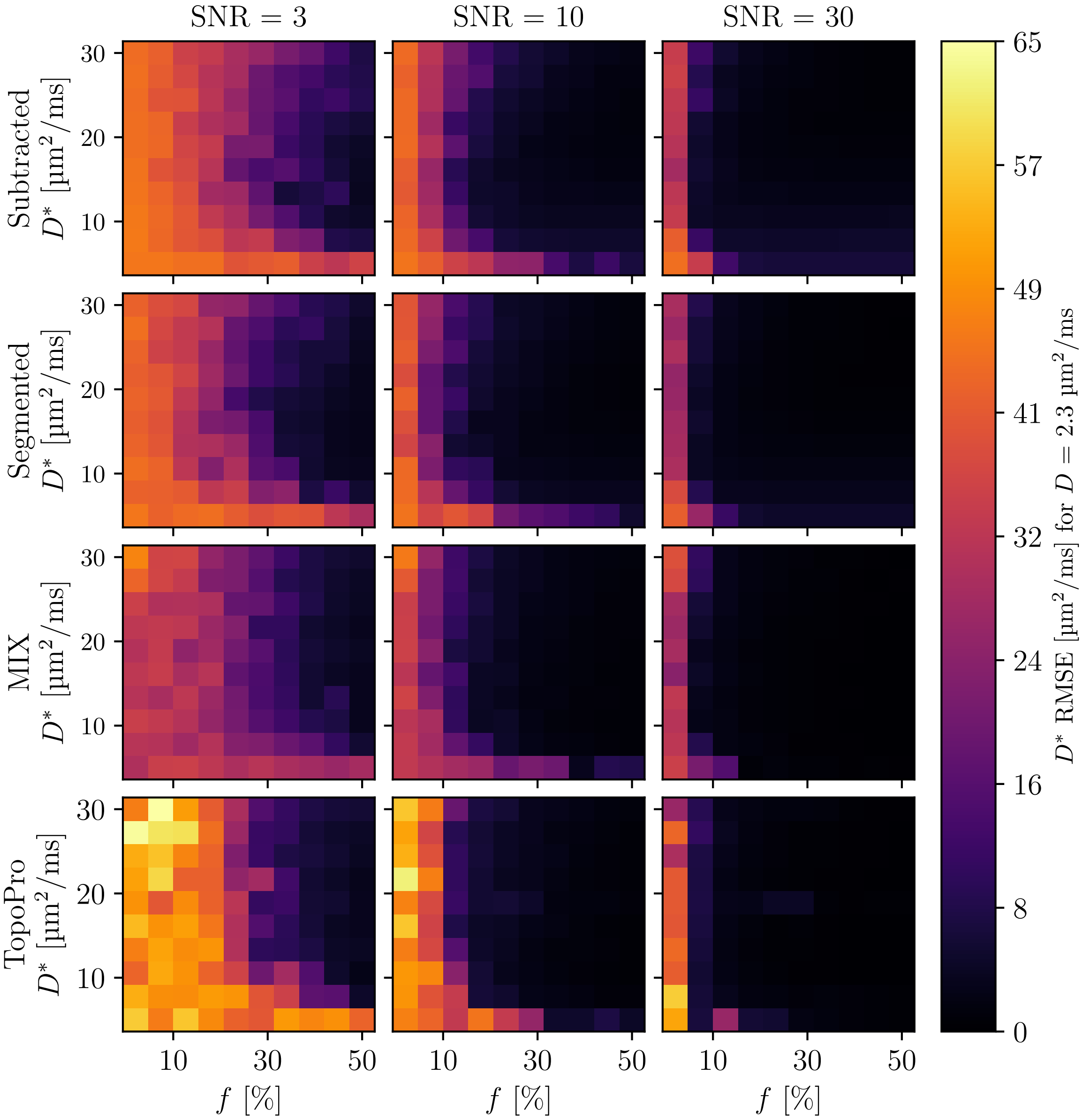

SNR = 3
SNR = 10
SNR = 30
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
10
30
50
f [%]
65
57
49
41
33
25
17
9
1
D^* RMSE [µm^2/ms] for $D = 0.7$ µm^2/ms

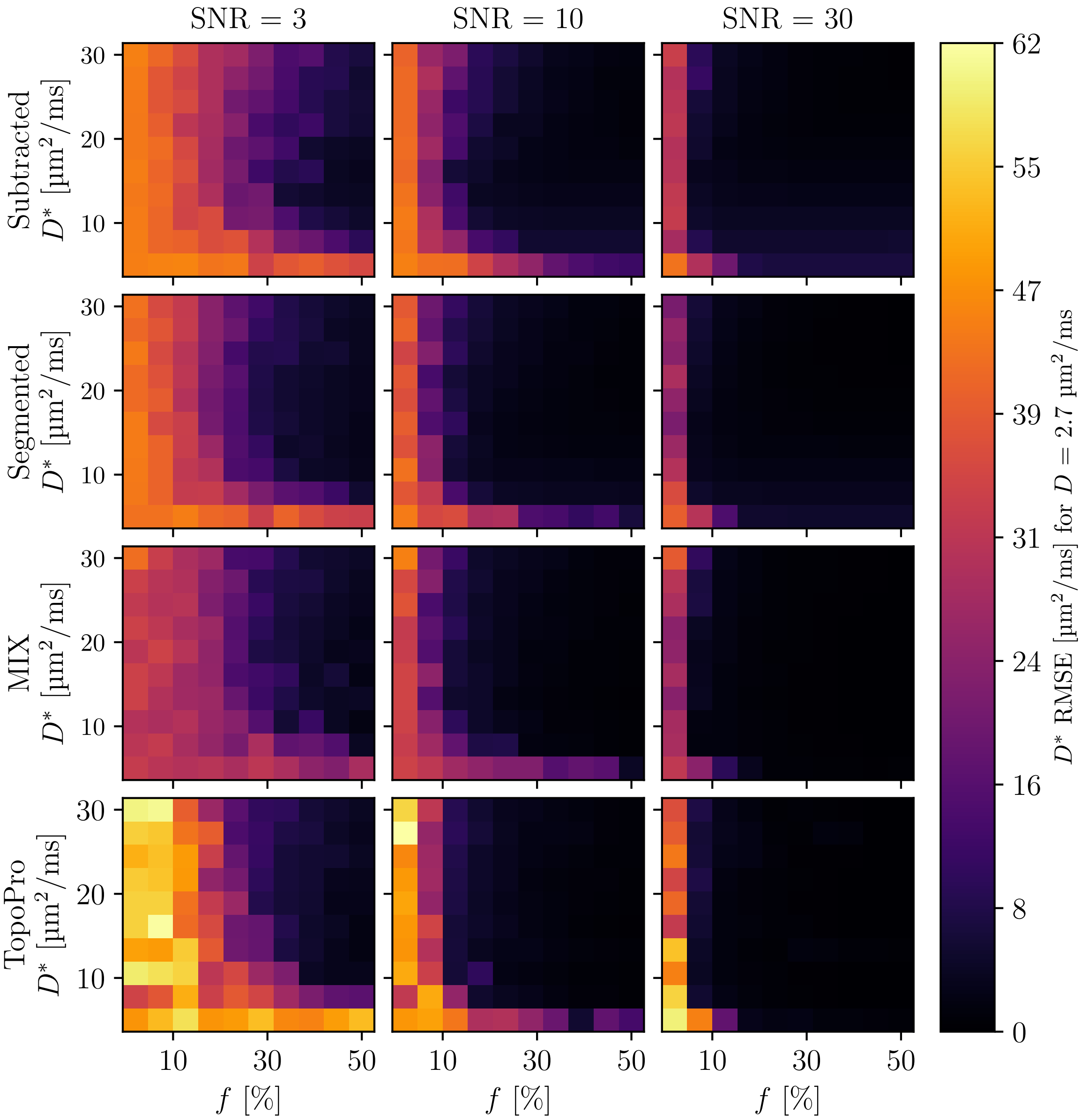

SNR = 3
SNR = 10
SNR = 30
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
10
30
50
f [%]
67
59
51
42
34
26
17
9
1
D^* RMSE [µm^2/ms] for D = 1.0 µm^2/ms

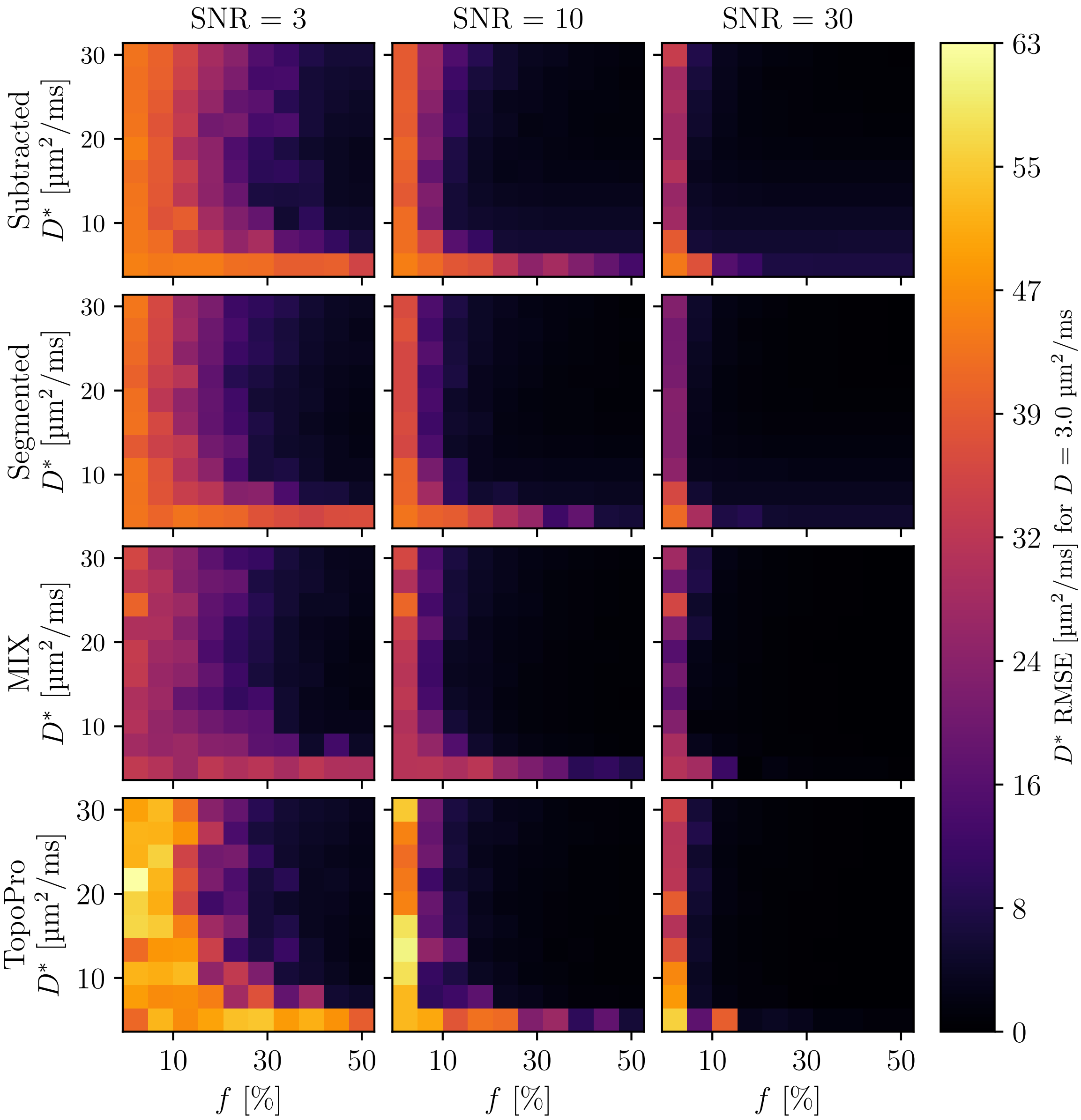

SNR = 3
SNR = 10
SNR = 30
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
10
30
50
f [%]
67
59
50
42
34
25
17
9
0
D^* RMSE [µm^2/ms] for $D = 1.4$ µm^2/ms


SNR = 3
SNR = 10
SNR = 30
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
30
20
10
10
30
50
f [%]
62
54
46
39
31
23
16
8
0
D^* RMSE [μm^2/ms] for $D = 1.7$ μm^2/ms


SNR = 3
SNR = 10
SNR = 30
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
10
20
30
10
30
50
f [%]
D^* RMSE [µm^2/ms] for D = 2.0 µm^2/ms
0
8
16
24
32
40
48
56
64

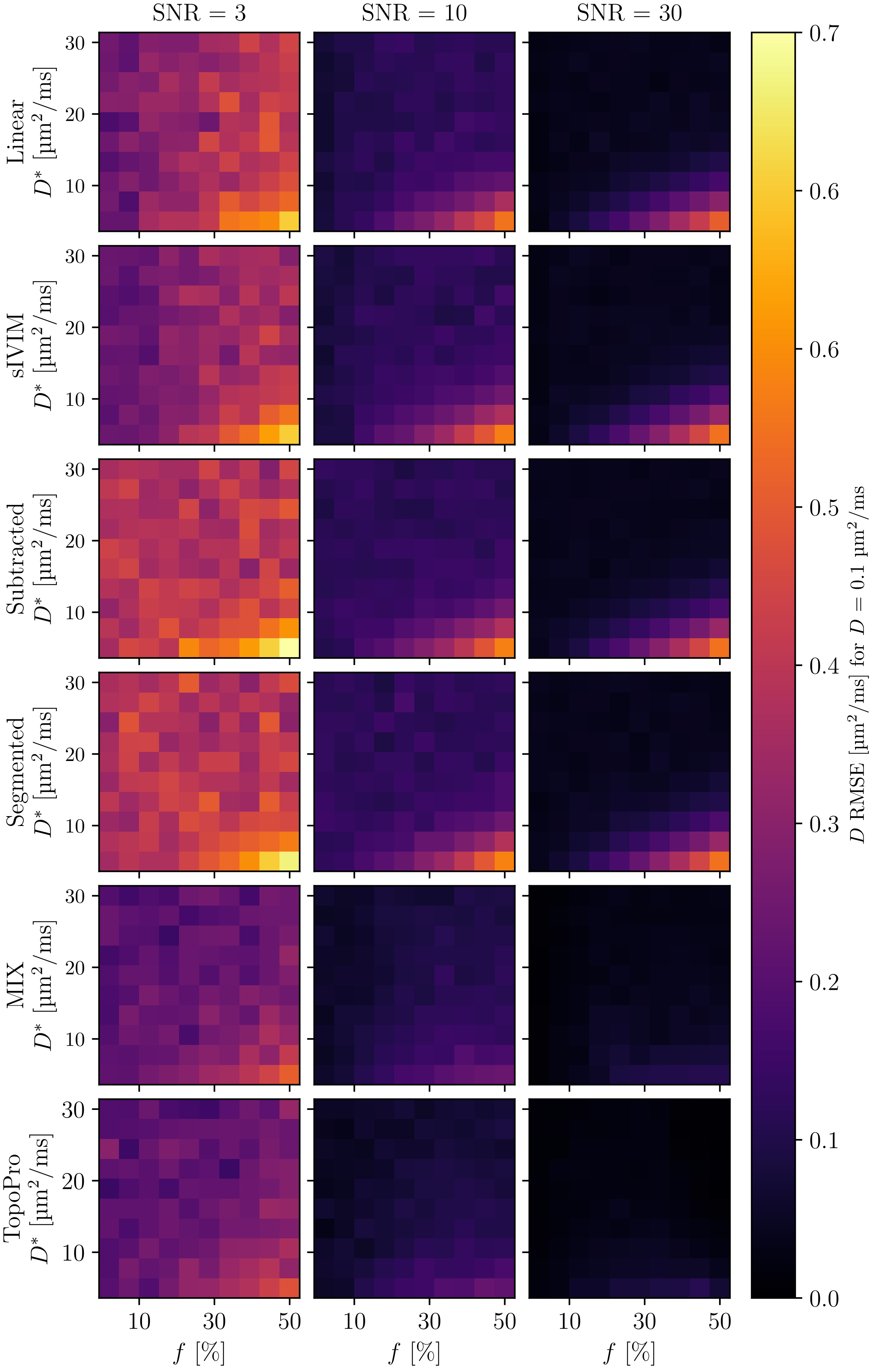

SNR = 3
SNR = 10
SNR = 30
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
10
20
30
f [%]
10
30
50
D^* RMSE [µm^2/ms] for $D = 2.3$ µm^2/ms
0
8
16
24
32
41
49
57
65

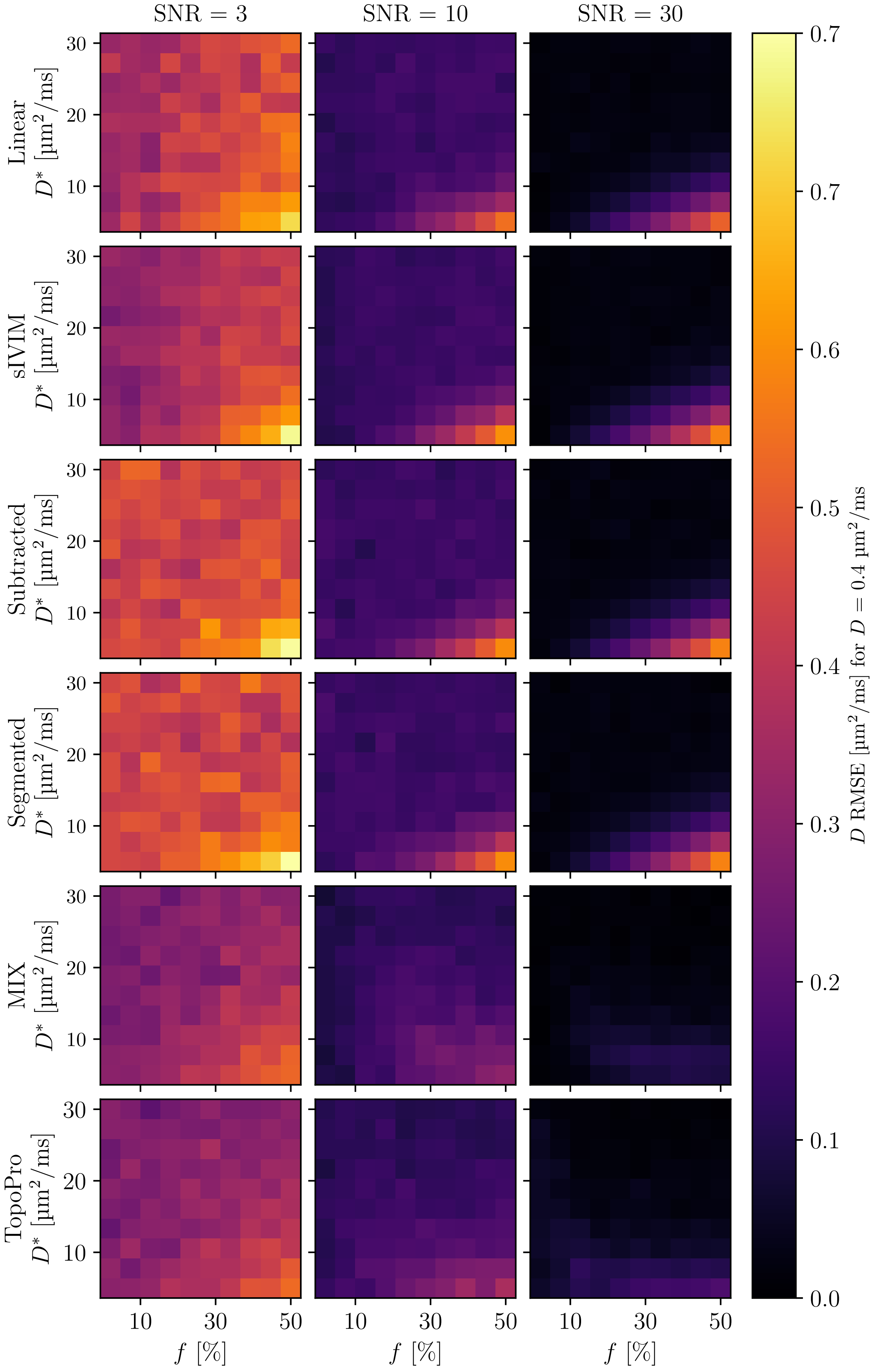

SNR = 3
SNR = 10
SNR = 30
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
10
30
50
f [%]
62
55
47
39
31
24
16
8
0
D^* RMSE [µm^2/ms] for D = 2.7 µm^2/ms

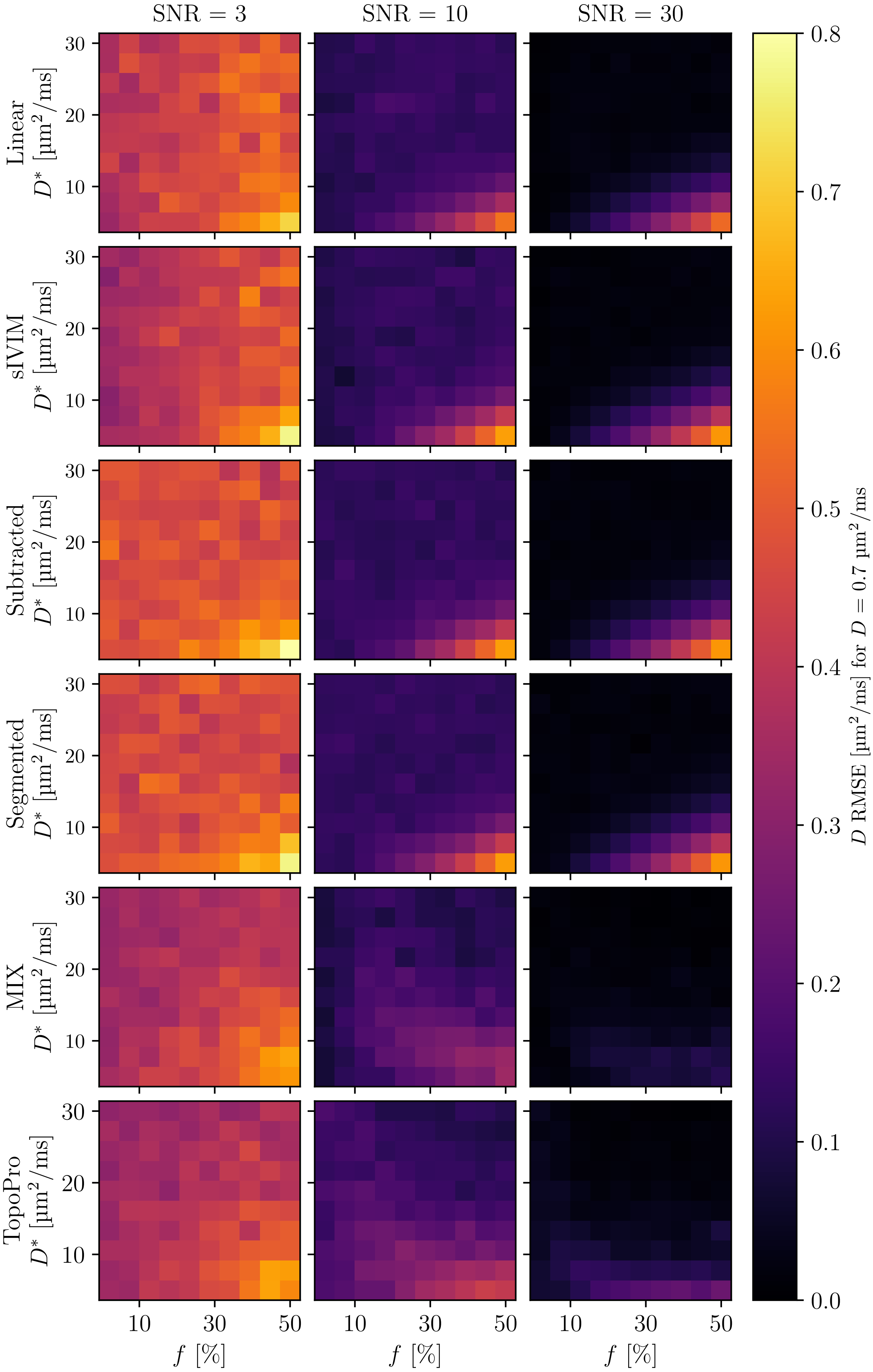


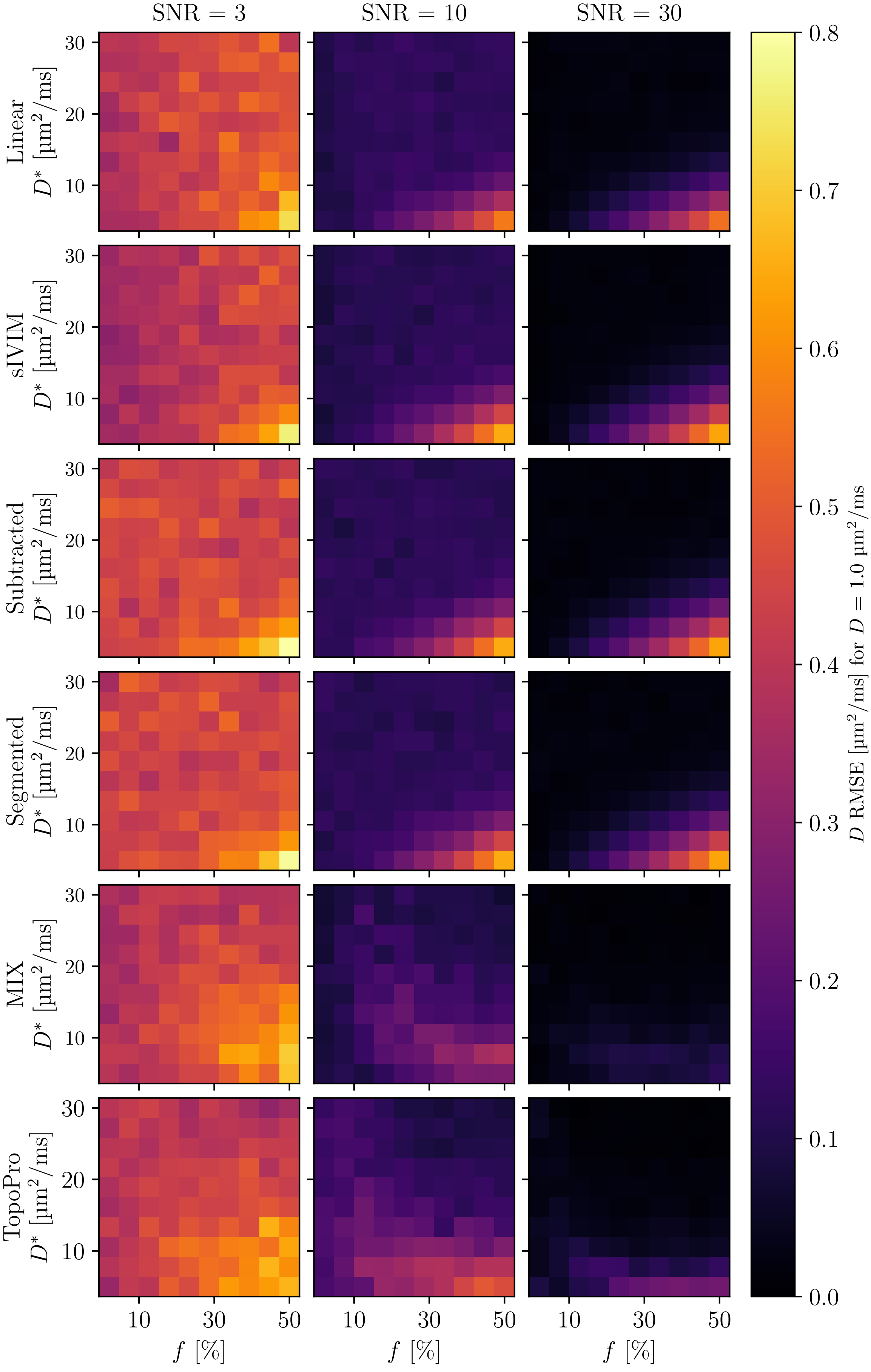
C.3 RMSE of D

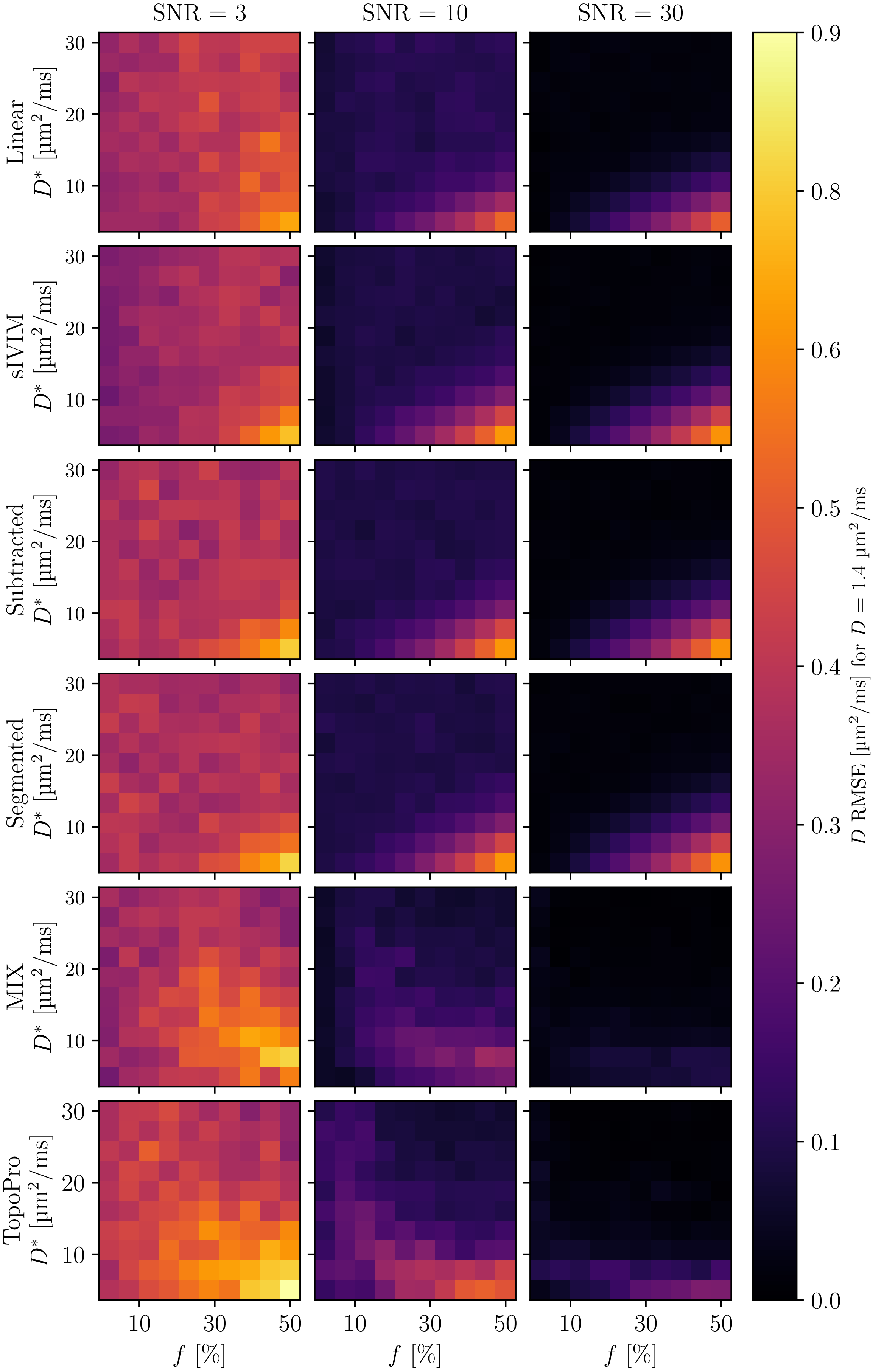
In this section, the RMSE maps of the D estimates for varying ground truth values of D in the slice direction are presented.


SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
30
20
10
f [%]
10
30
50
D RMSE [μm^2/ms] for $D = 0.1$ μm^2/ms
0.7
0.6
0.6
0.5
0.4
0.3
0.2
0.1
0.0

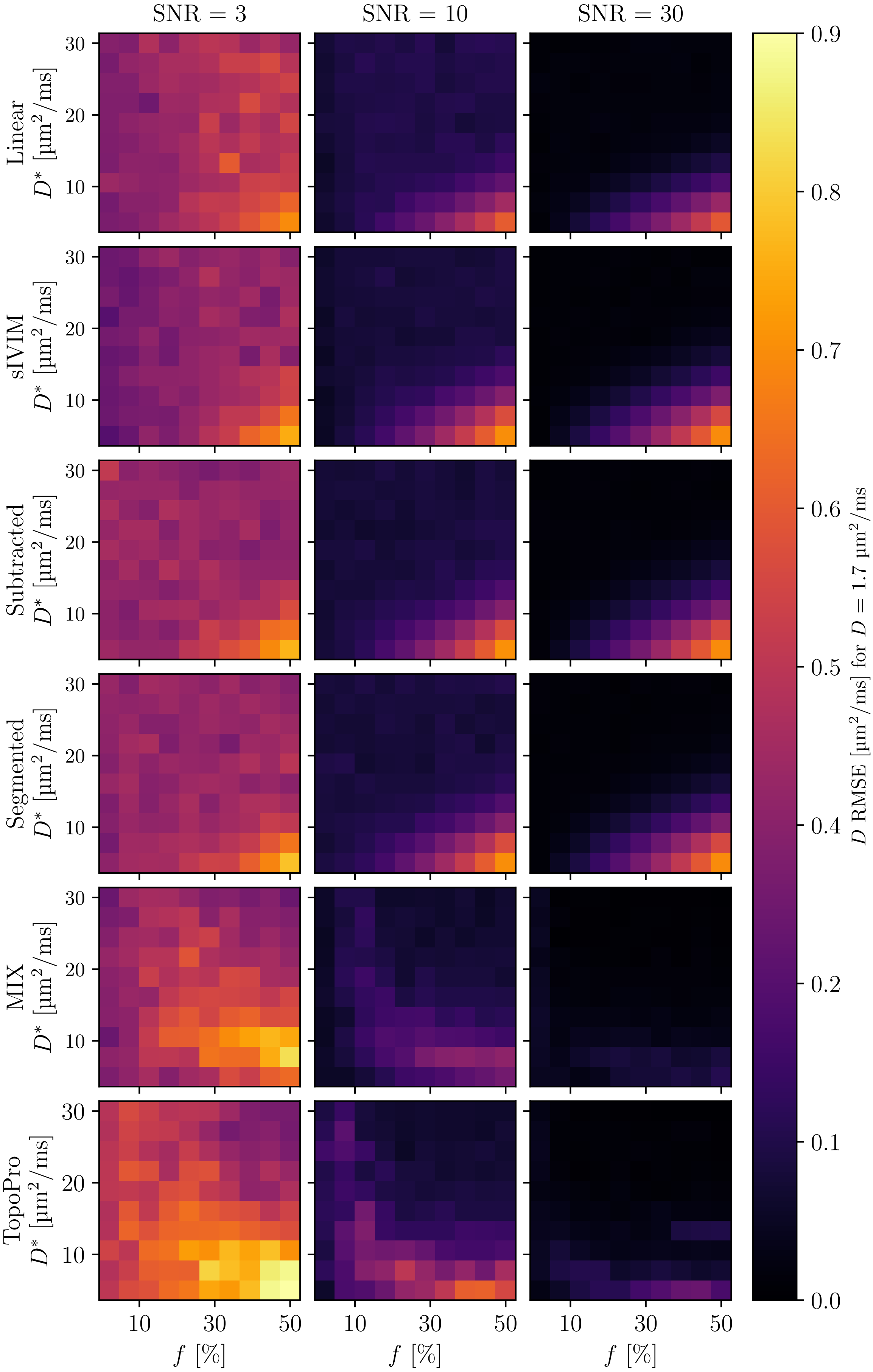

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
10
20
30
f [%]
10
30
50
0.7
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
D RMSE [µm^2/ms] for $D = 0.4$ µm^2/ms

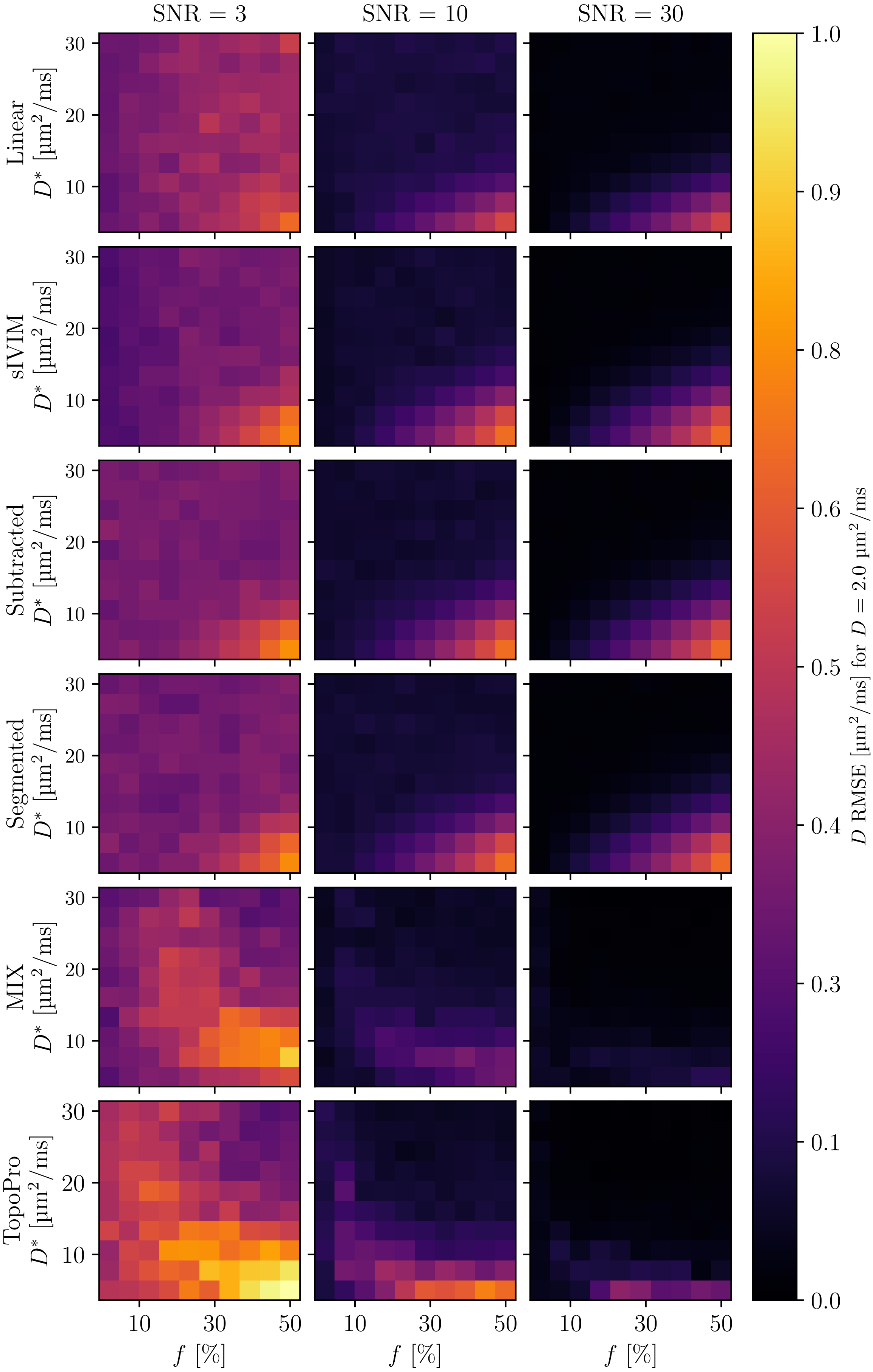

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
10
30
50
f [%]
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
D RMSE [µm^2/ms] for D = 0.7 µm^2/ms

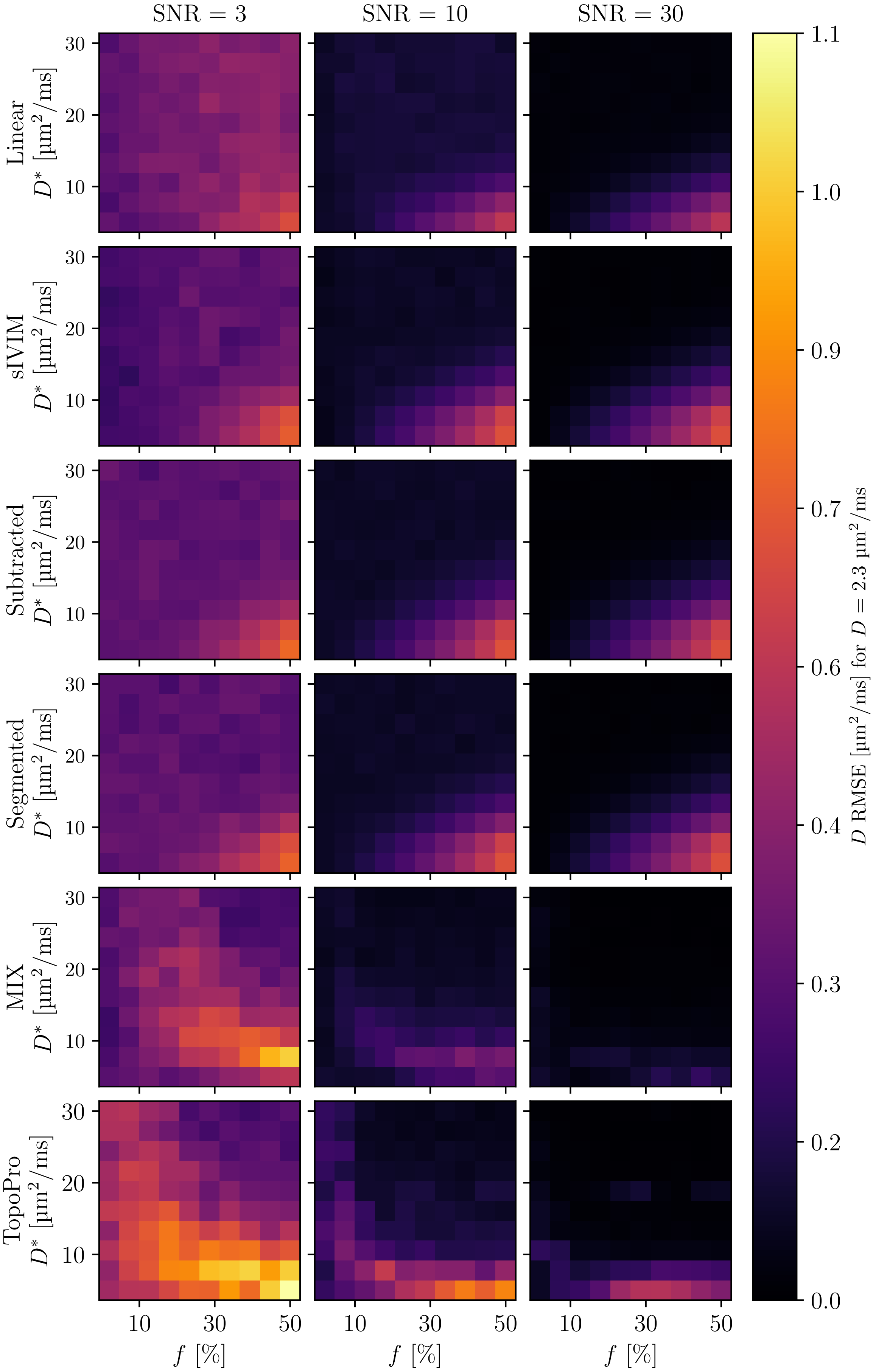

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
10
30
50
f [%]
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
D RMSE [µm^2/ms] for $D = 1.0$ µm^2/ms

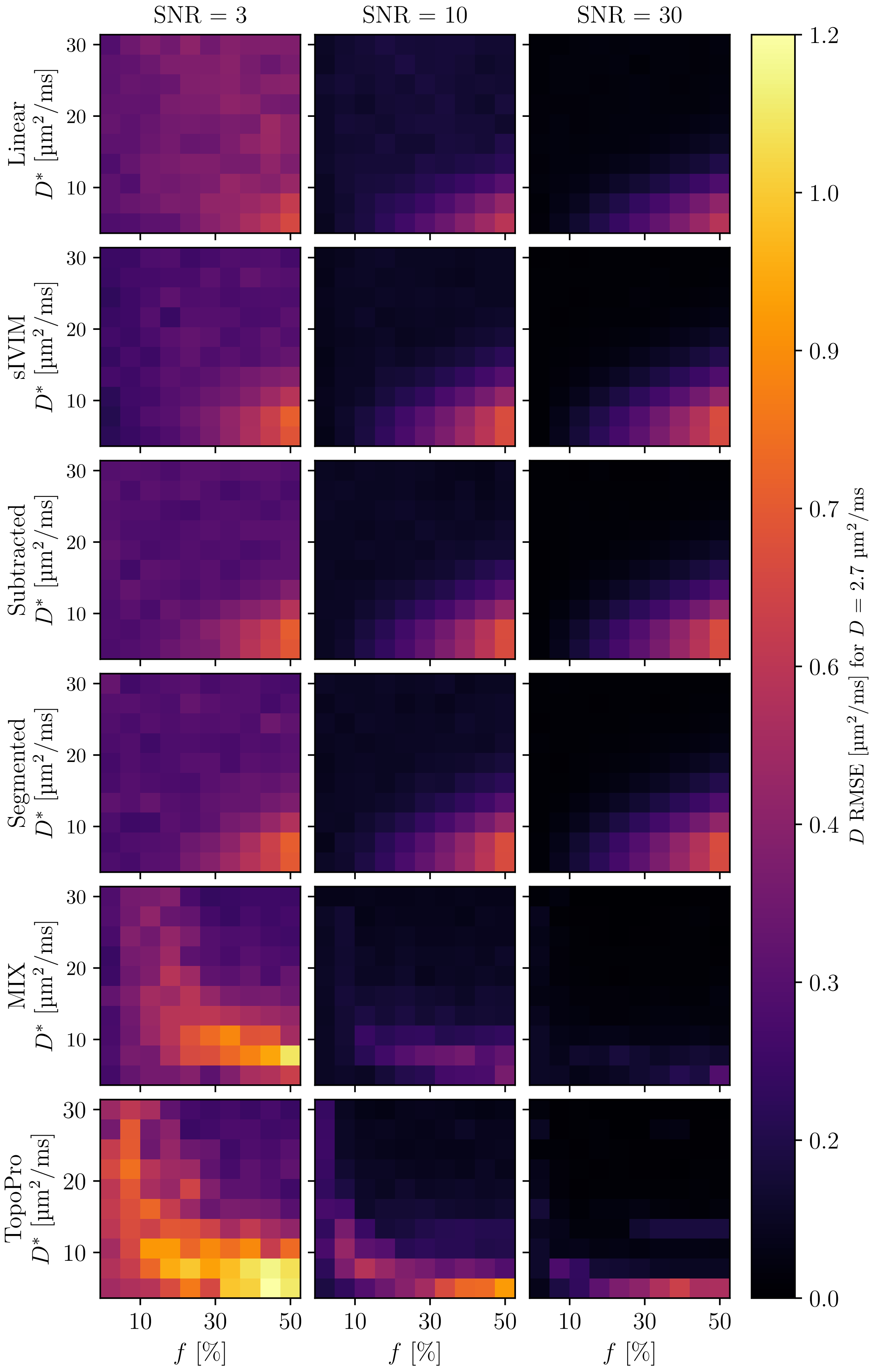


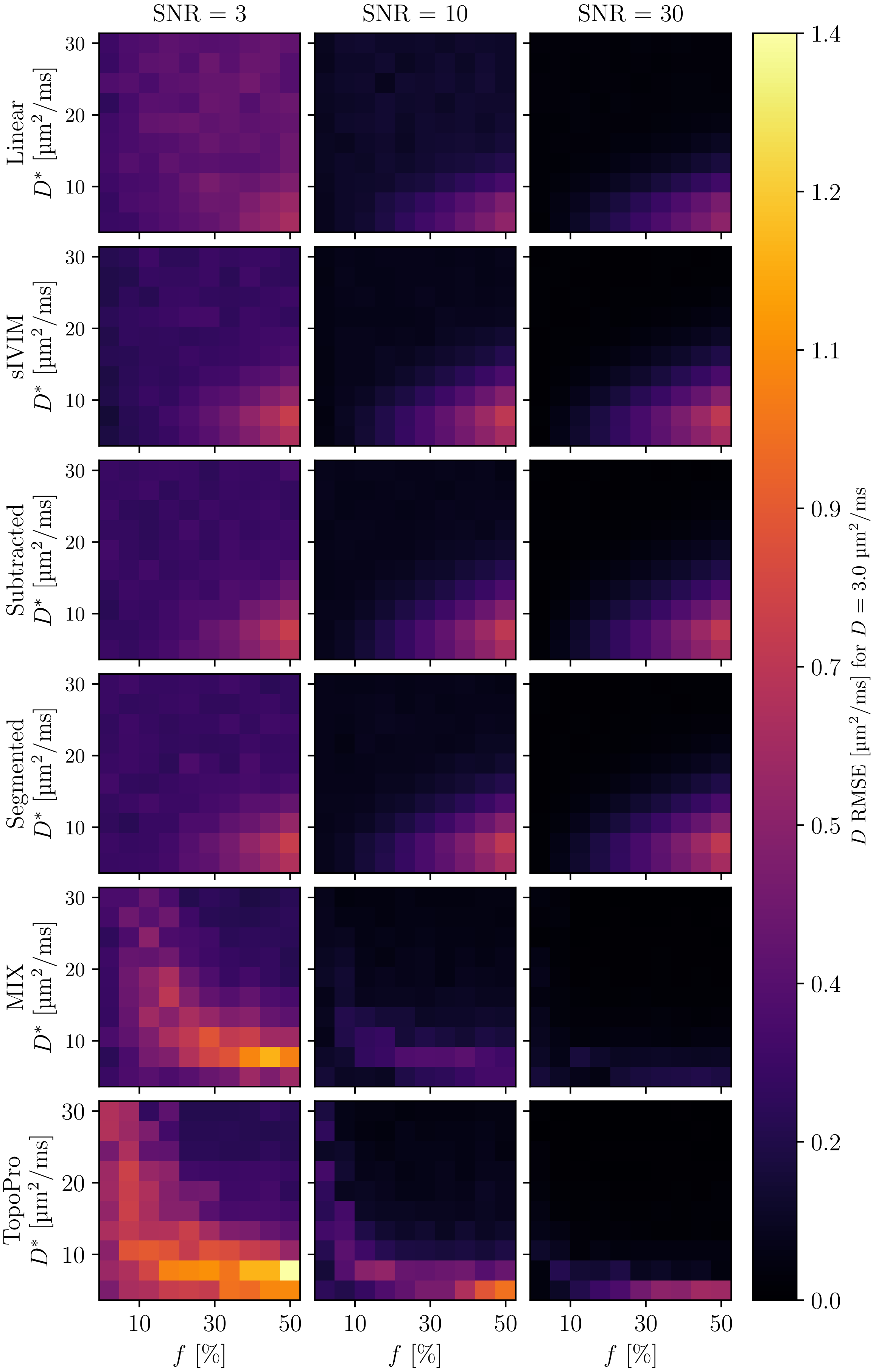
SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
f [%]
10
30
50
D RMSE [µm^2/ms] for $D = 1.4$ µm^2/ms
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
30
20
10
f [%]
10
30
50
D RMSE [µm^2/ms] for D = 1.7 µm^2/ms
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
10
20
30
10
30
50
f [%]
D RMSE [µm^2/ms] for $D = 2.0$ µm^2/ms
0.0
0.1
0.3
0.4
0.5
0.6
0.8
0.9
1.0


SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
30
20
10
10
30
50
f [%]
D RMSE [μm^2/ms] for D = 2.3 μm^2/ms
1.1
1.0
0.9
0.7
0.6
0.4
0.3
0.2
0.0

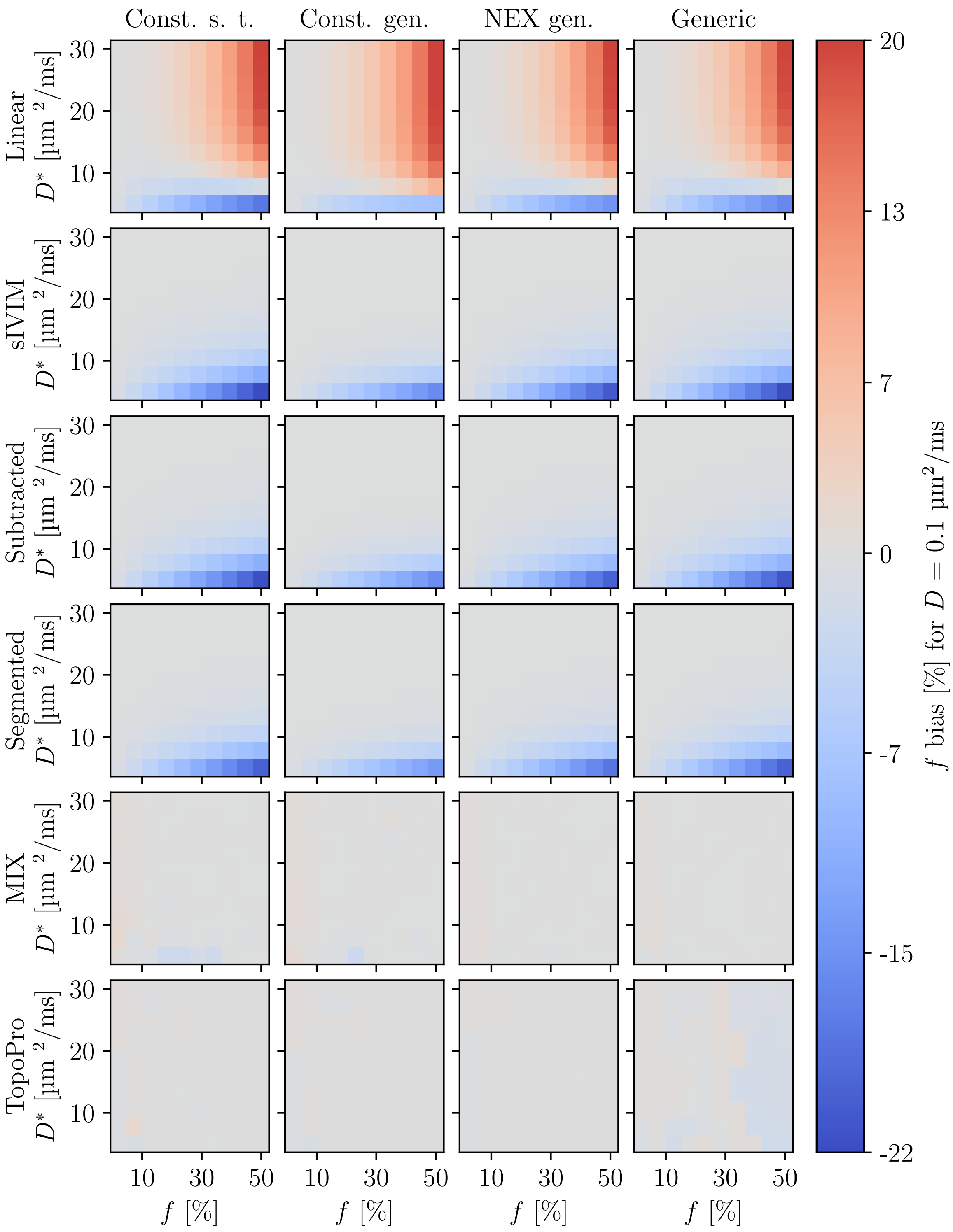

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm^2/ms]
10
20
30
f [%]
10
30
50
D RMSE [µm^2/ms] for D = 2.7 µm^2/ms
0.0
0.2
0.3
0.4
0.6
0.7
0.9
1.0
1.2

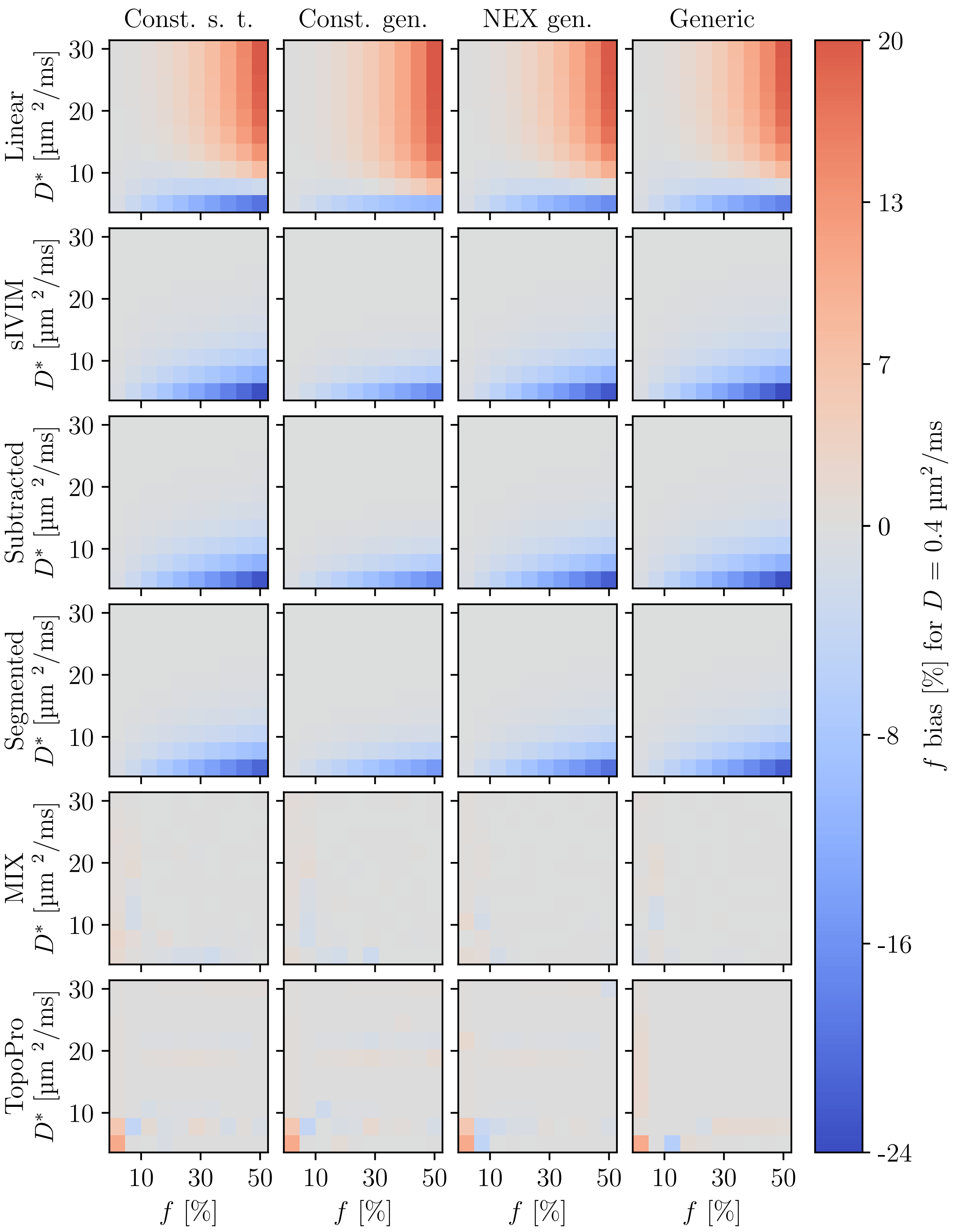

SNR = 3
SNR = 10
SNR = 30
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm^2/ms]
10
20
30
f [%]
10
30
50
0.0
0.2
0.4
0.5
0.7
0.9
1.1
1.2
1.4
D RMSE [μm^2/ms] for D = 3.0 μm^2/ms

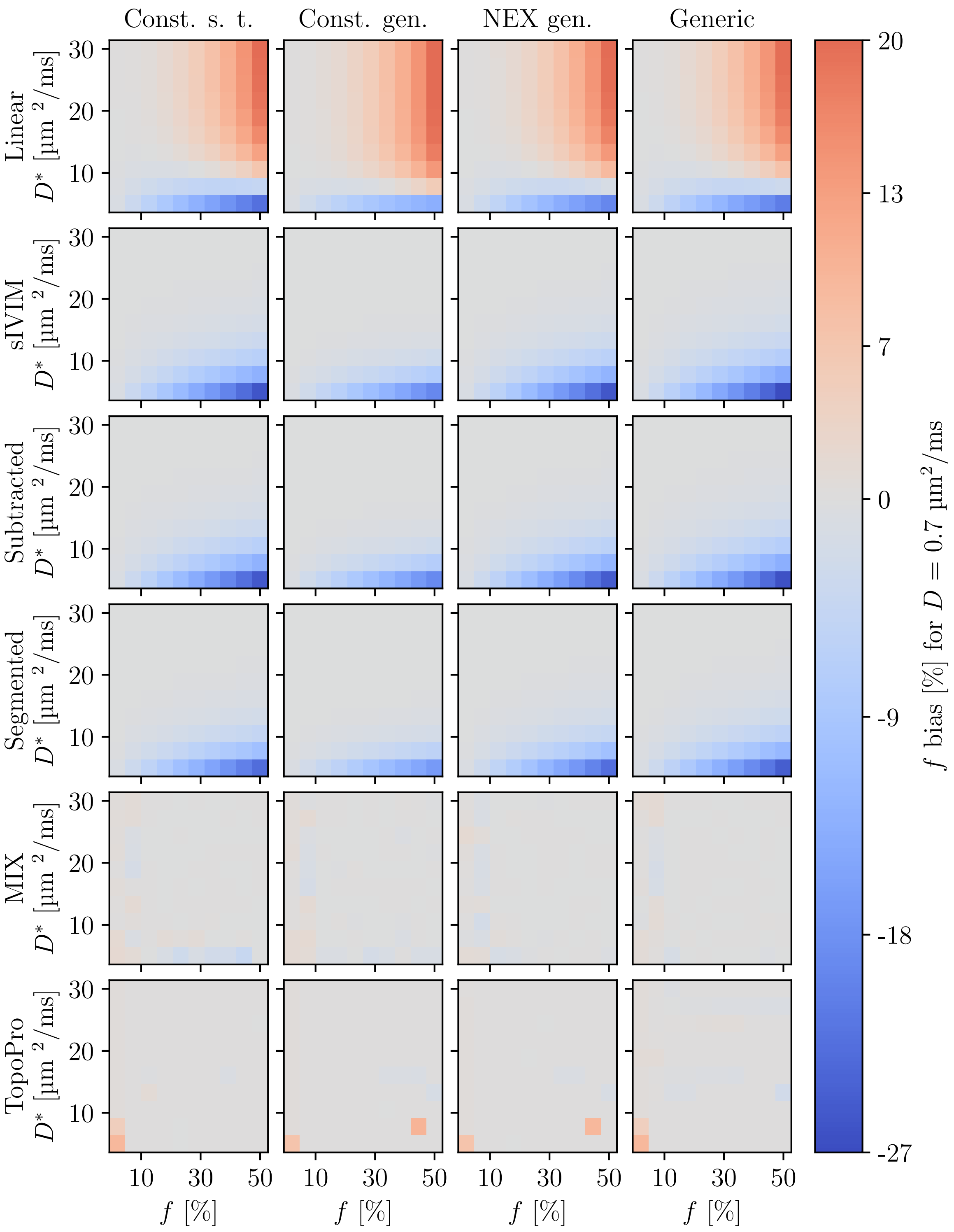
D Appendix: Bias Maps

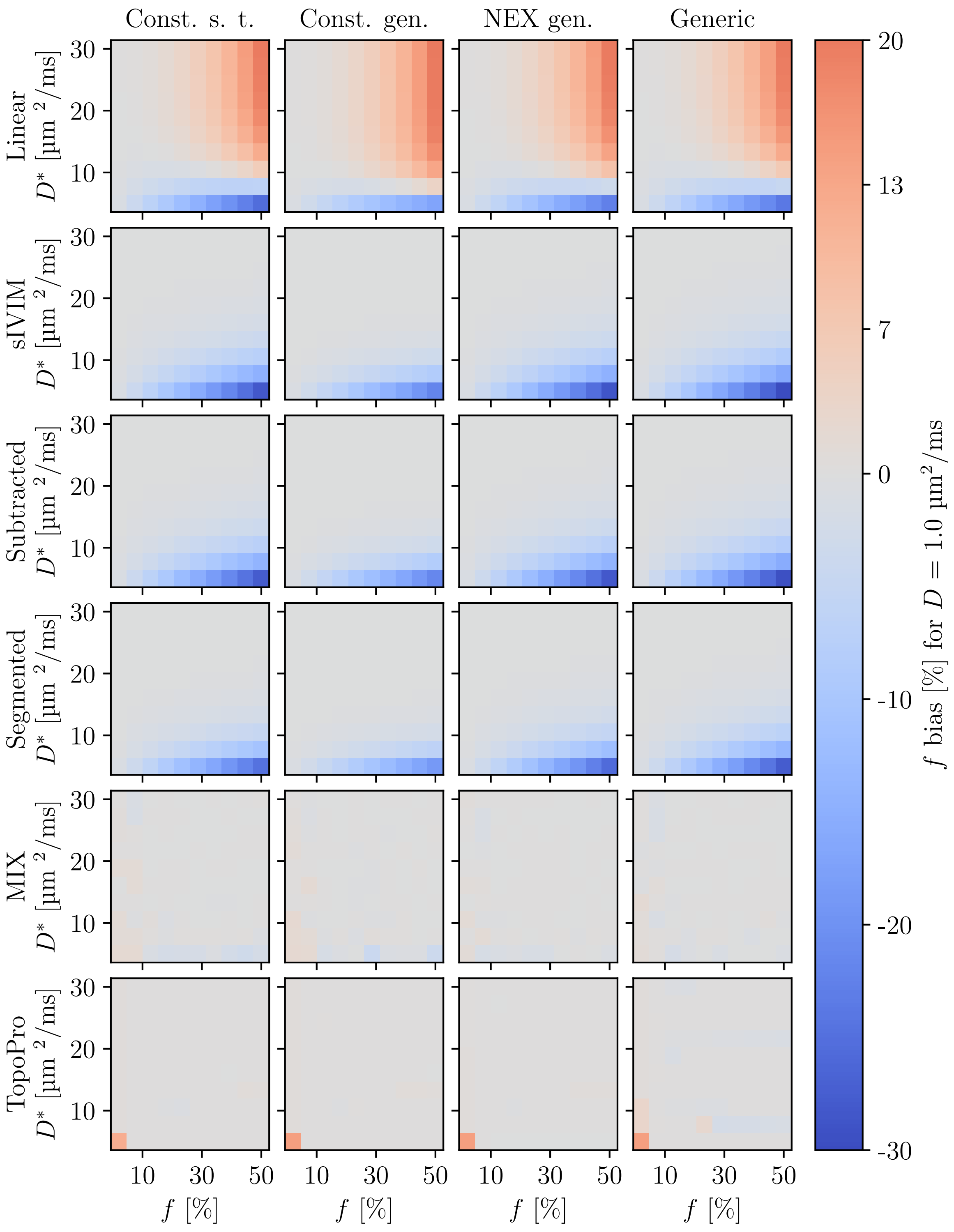
D.1 Bias in f

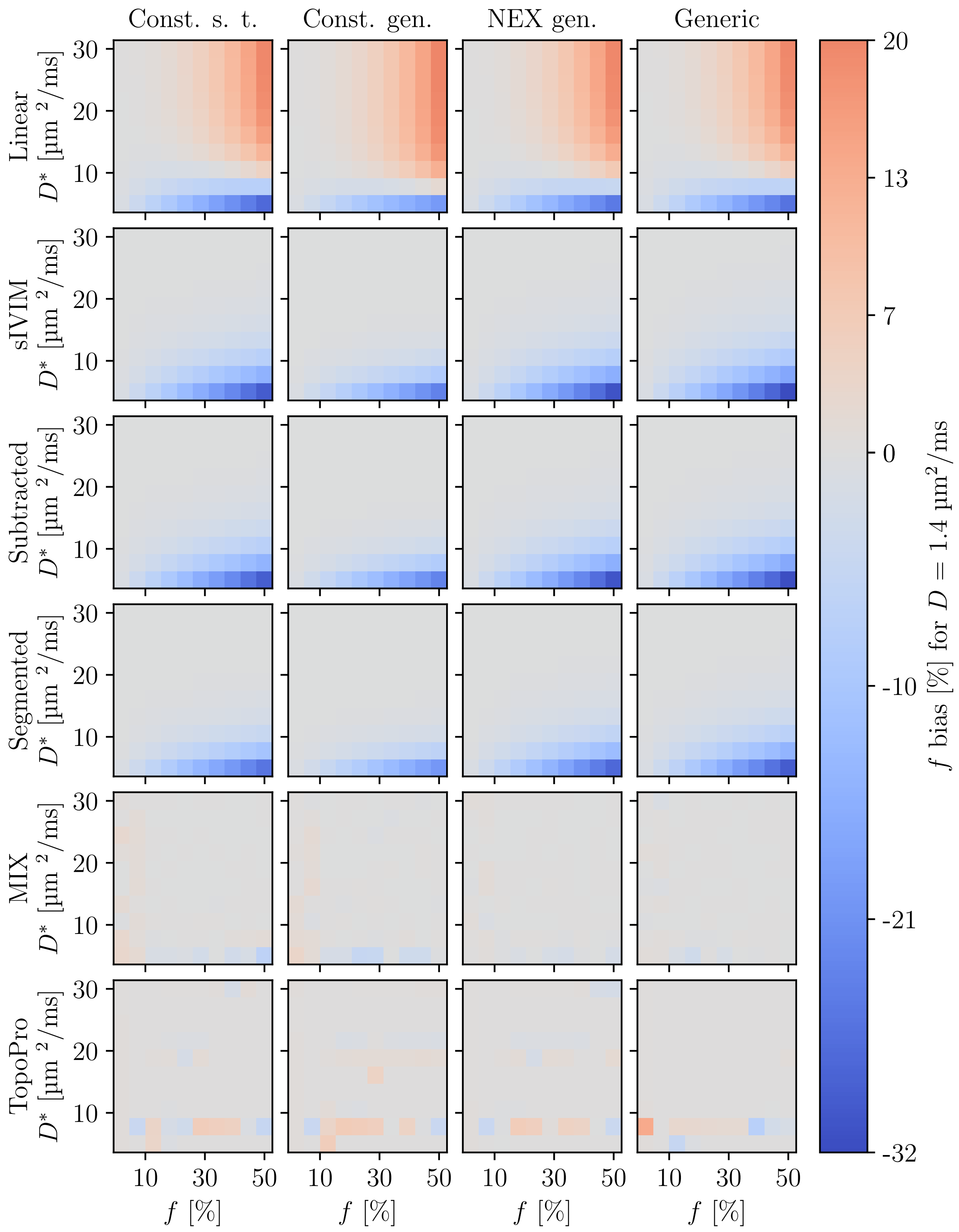
In this section, bias maps for the f estimates are presented for all tested protocols, with varying ground truth values of D in the slice direction. The bias is presented in absolute units, i.e. the percentage unit of the maps should not be interpreted as a relative error. Note that the scales for the positive (red) and negative (blue) colours differ from each other.

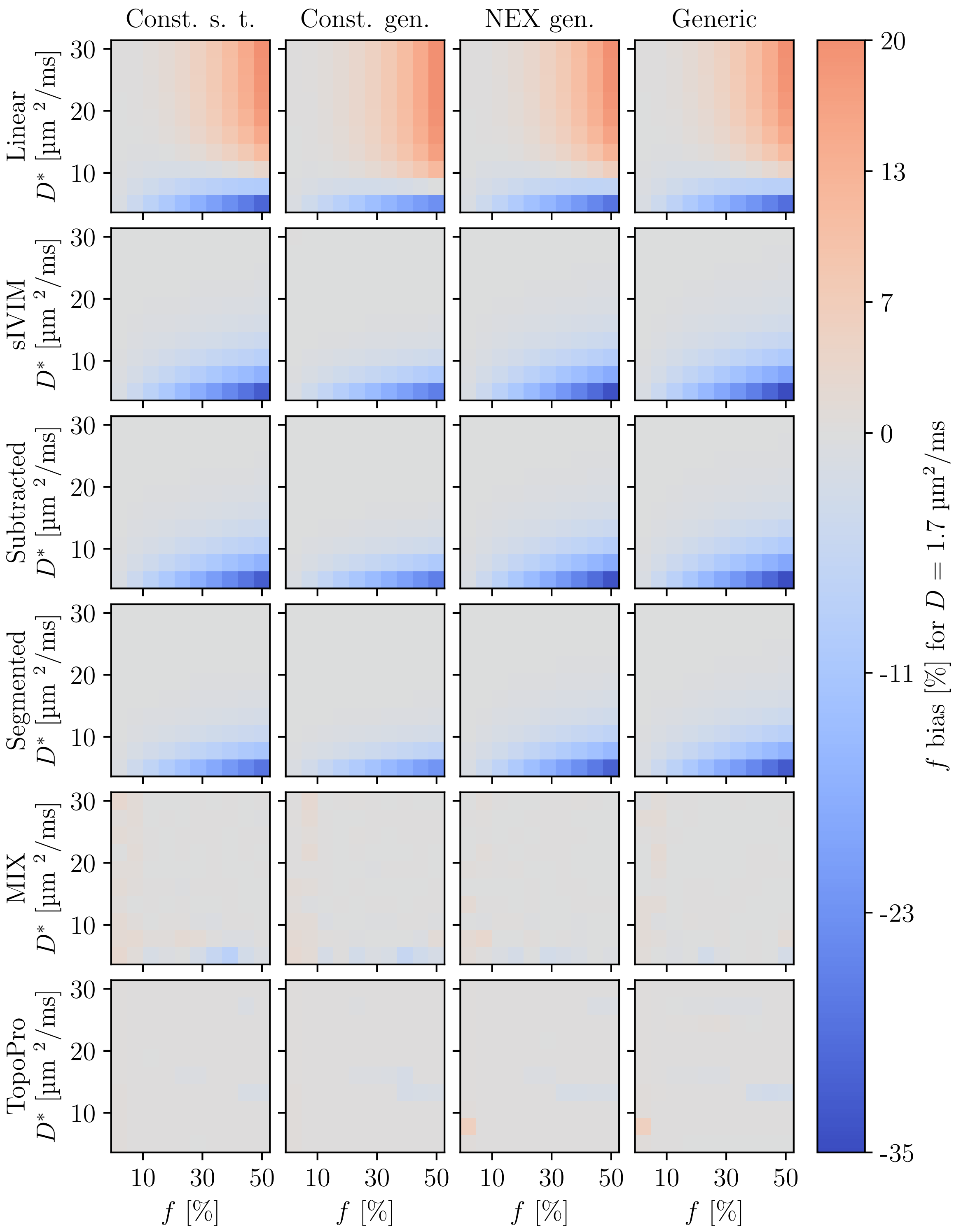

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
30
20
10
10
30
50
f [%]
20
13
7
0
-7
-15
-22
f bias [%] for $D = 0.1$ μm^2/ms

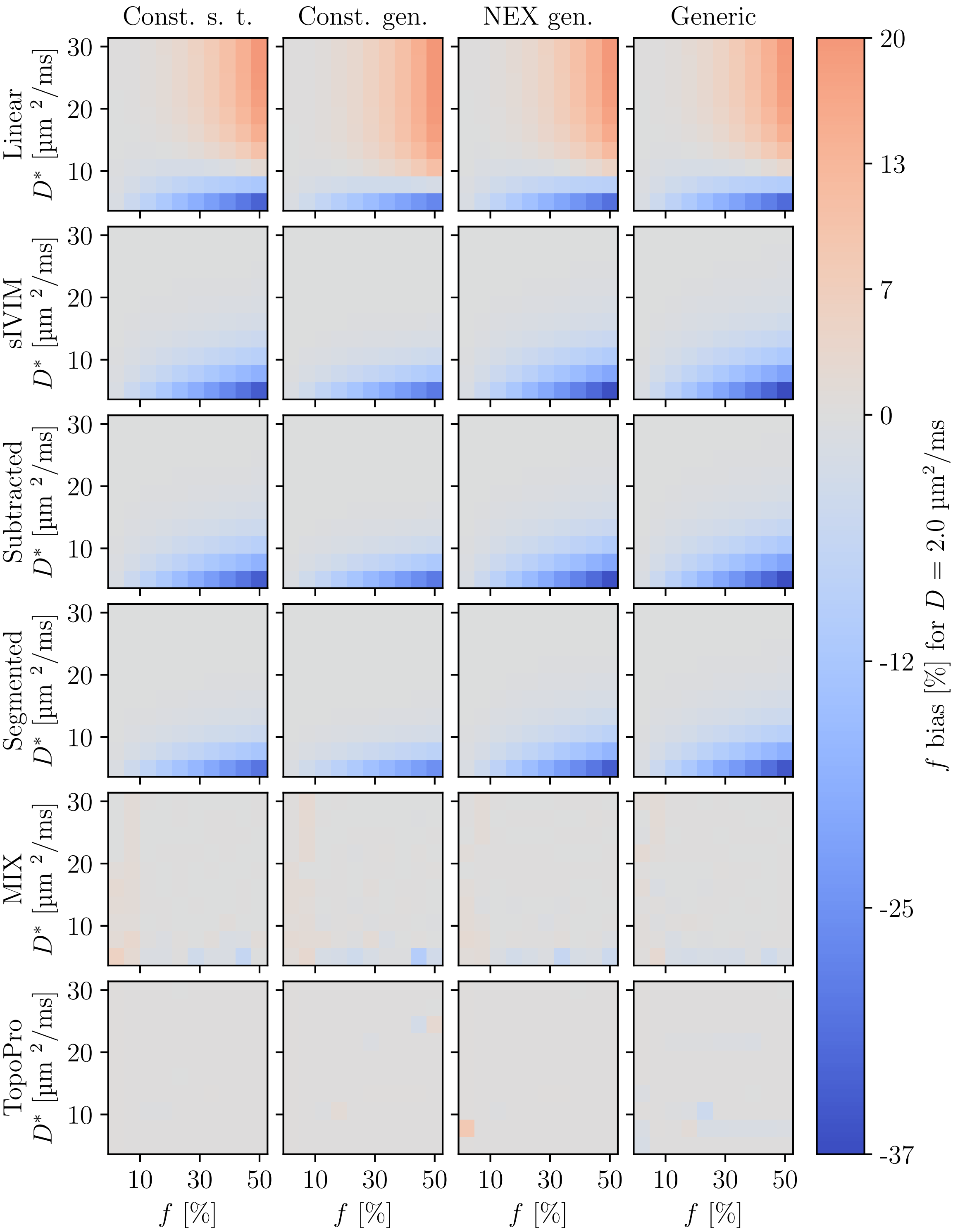

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
30
20
10
10
30
50
f [%]
20
13
7
0
-8
-16
-24
f bias [%] for $D = 0.4$ μm^2/ms

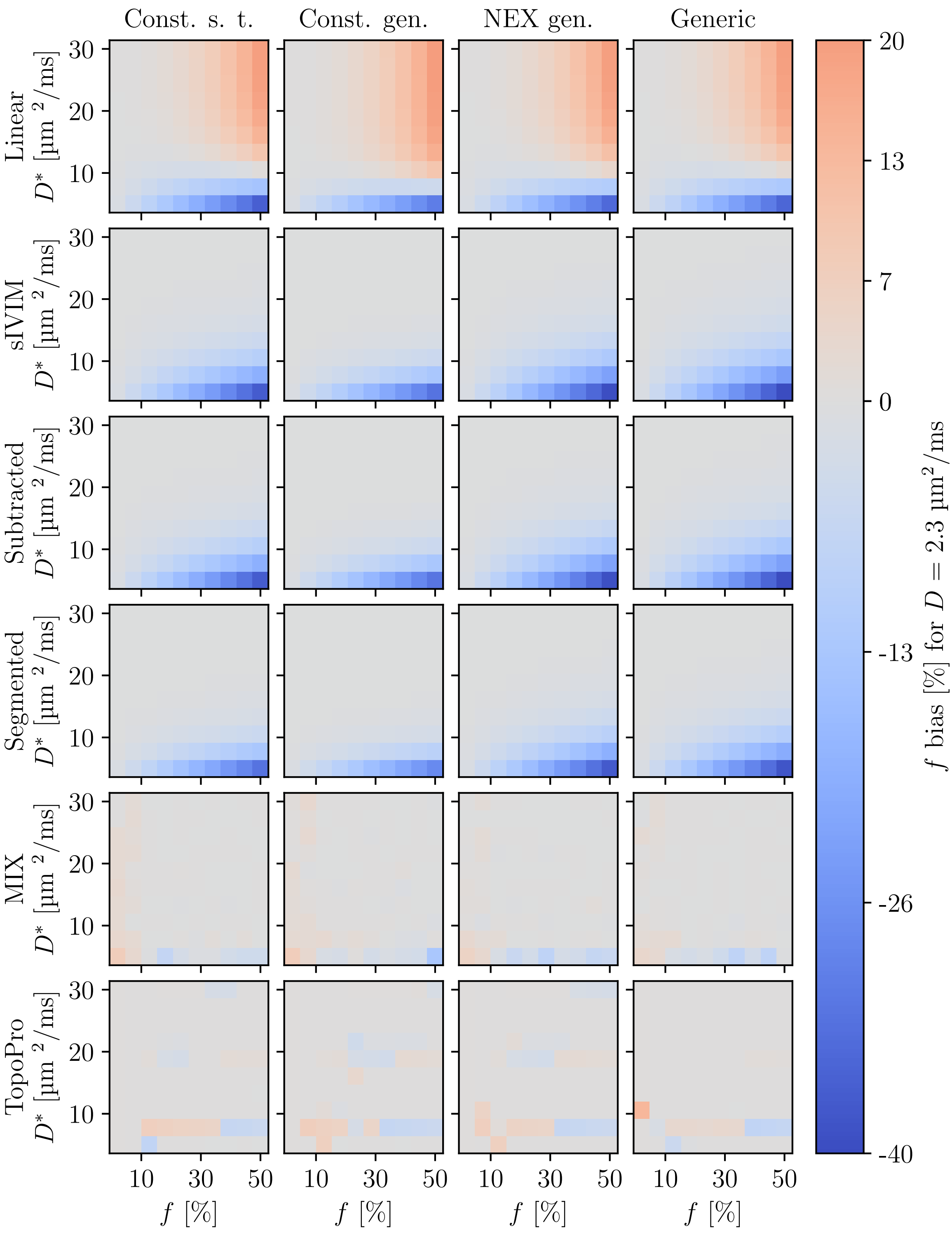

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
10
20
30
10
30
50
f [%]
20
13
7
0
-9
-18
-27
f bias [%] for D = 0.7 μm^2/ms

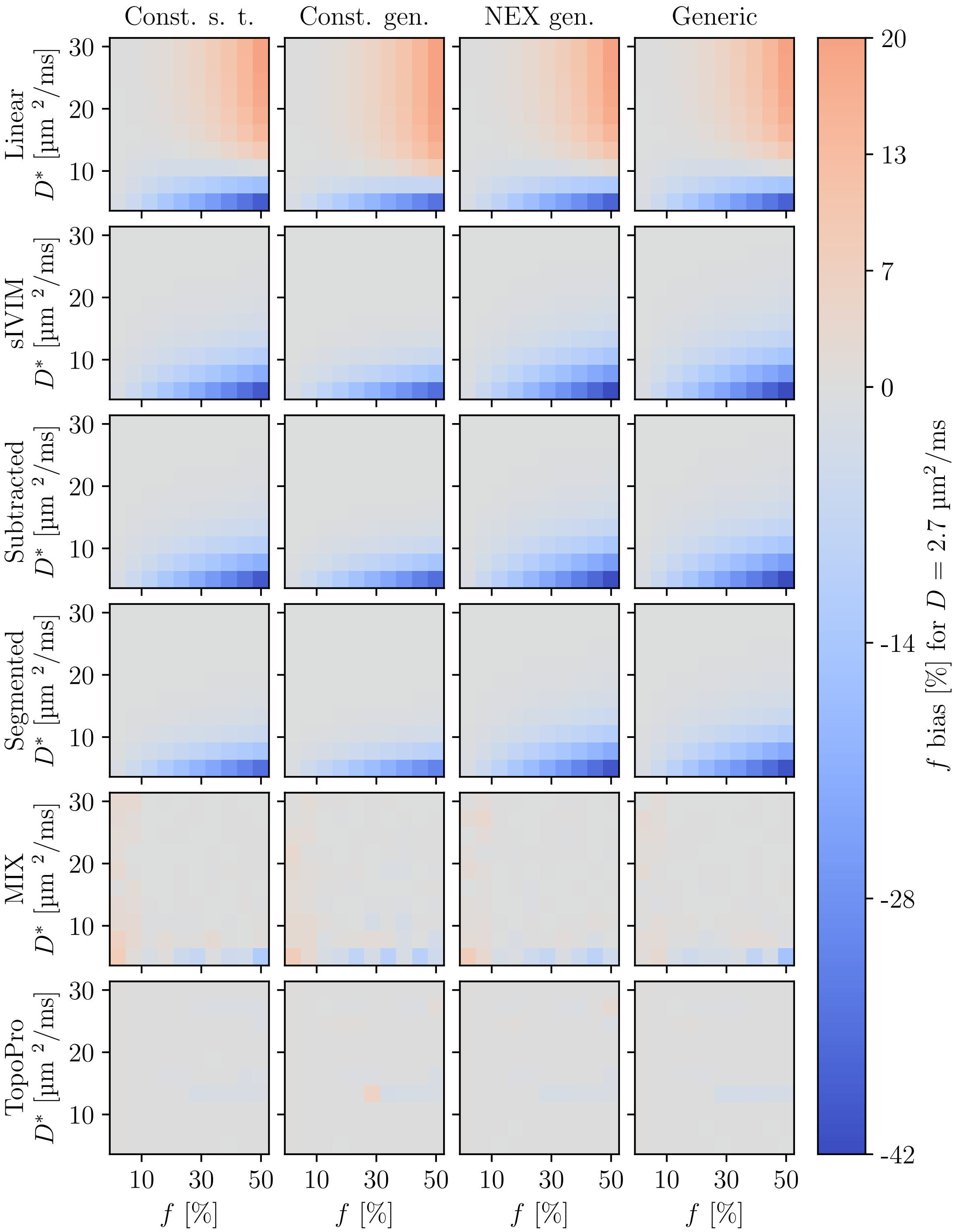

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D* [μm 2/ms]
30
20
10
10
30
50
f [%]
20
13
7
0
-10
-20
-30
f bias [%] for D = 1.0 μm²/ms

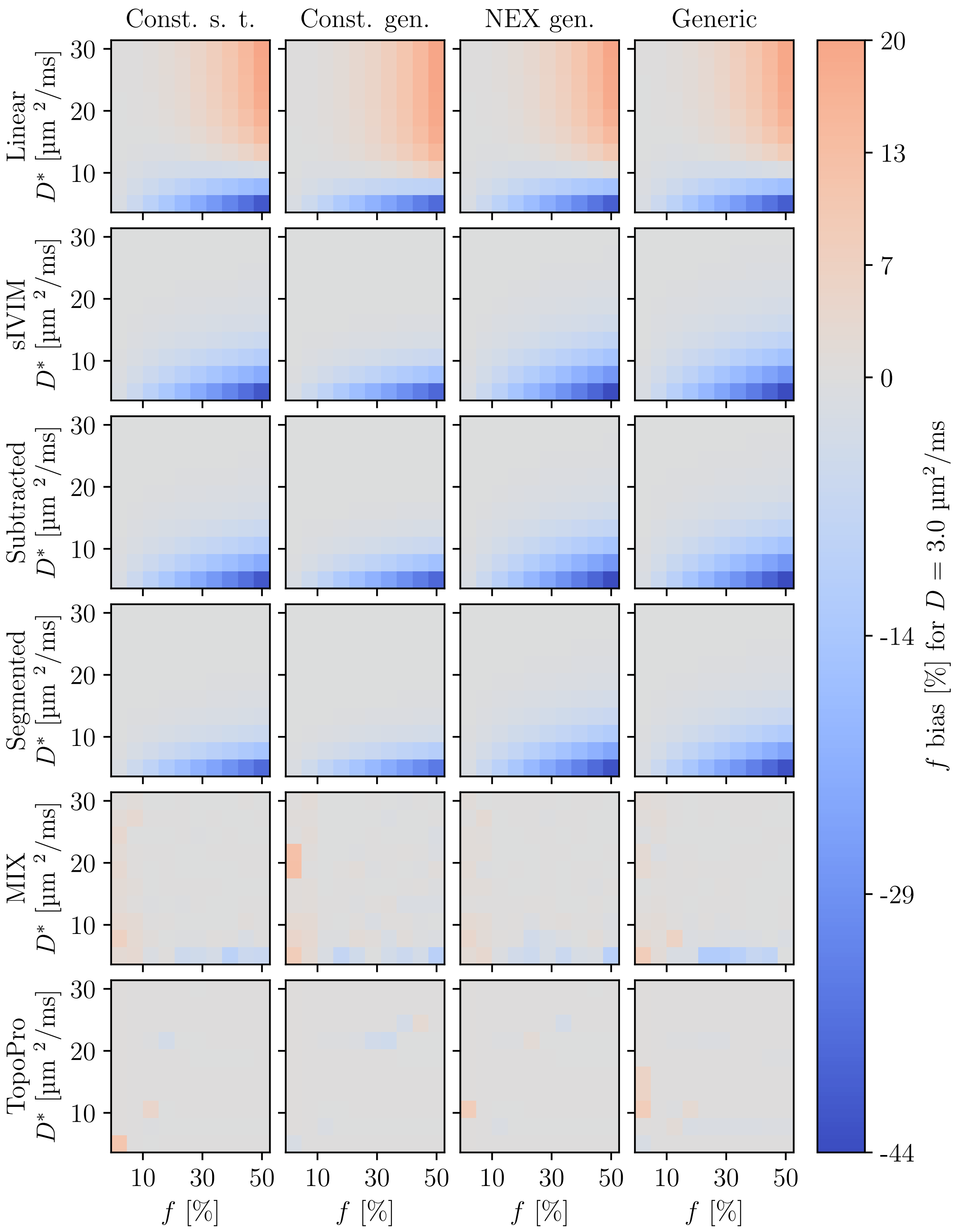

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
10
20
30
10
30
50
f [%]
20
13
7
0
-10
-21
-32
f bias [%] for D = 1.4 µm^2/ms


Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
D^* [µm ^{2}/ms]
sIVIM
D^* [µm ^{2}/ms]
Subtracted
D^* [µm ^{2}/ms]
Segmented
D^* [µm ^{2}/ms]
MIX
D^* [µm ^{2}/ms]
TopoPro
D^* [µm ^{2}/ms]
10
20
30
10
30
50
f [%]
f bias [%] for D = 1.7 µm^2/ms
20
13
7
0
-11
-23
-35


Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
20
13
7
0
-12
-25
-37
f bias [%] for $D = 2.0$ µm^2/ms

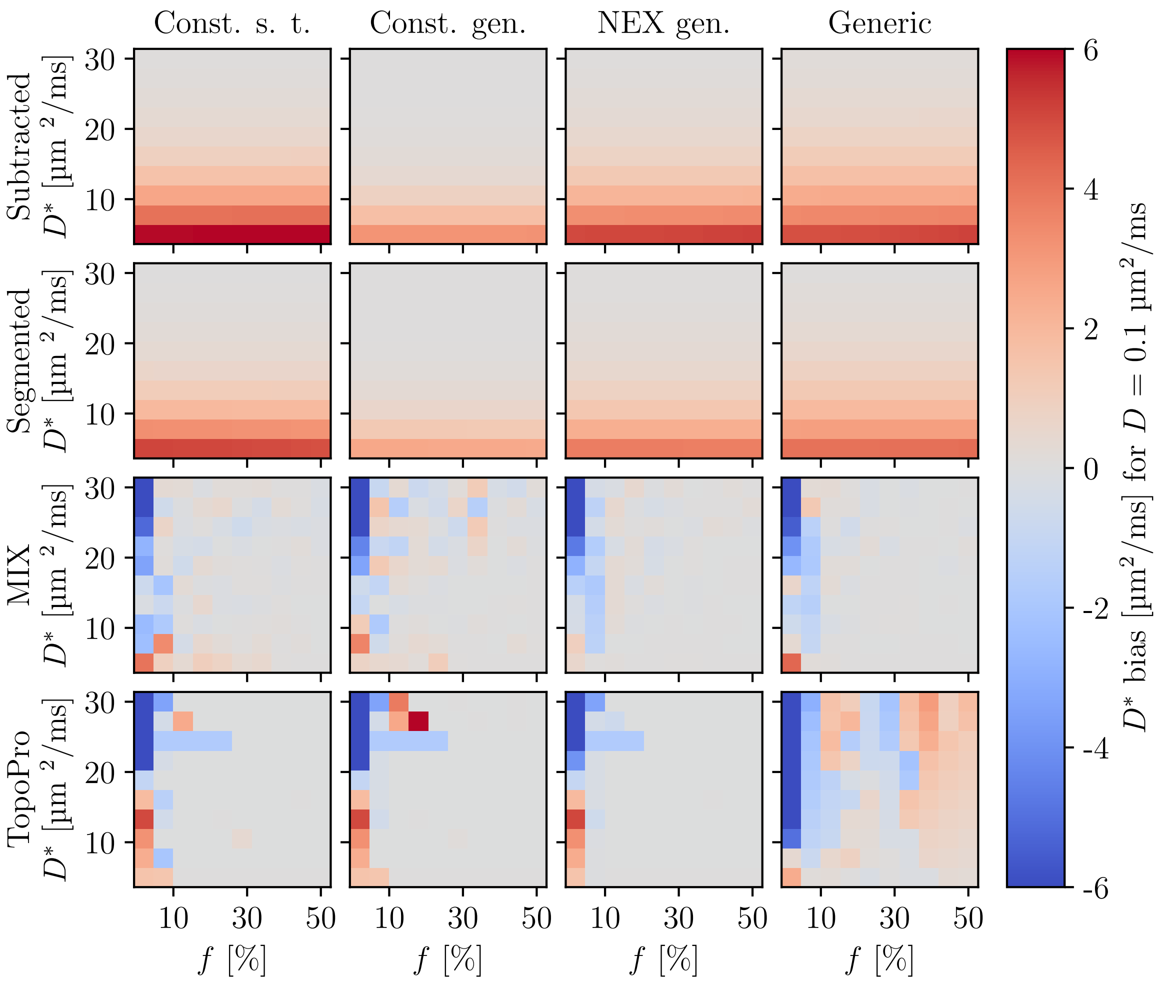

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
10
20
30
10
30
50
f [%]
20
13
7
0
-13
-26
-40
f bias [%] for $D = 2.3$ μm^2/ms

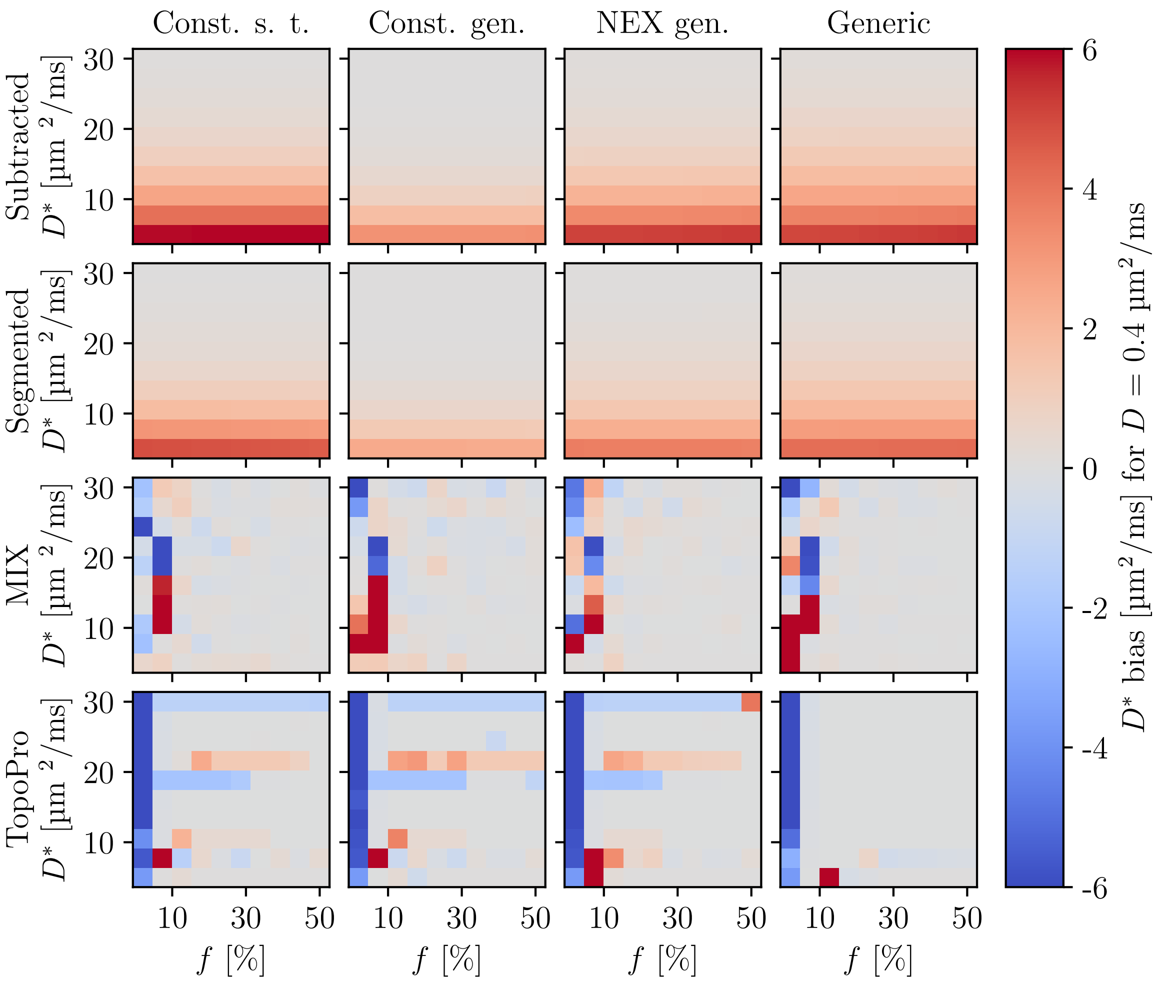

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
20
13
7
0
-14
-28
-42
f bias [%] for $D = 2.7$ µm^2/ms

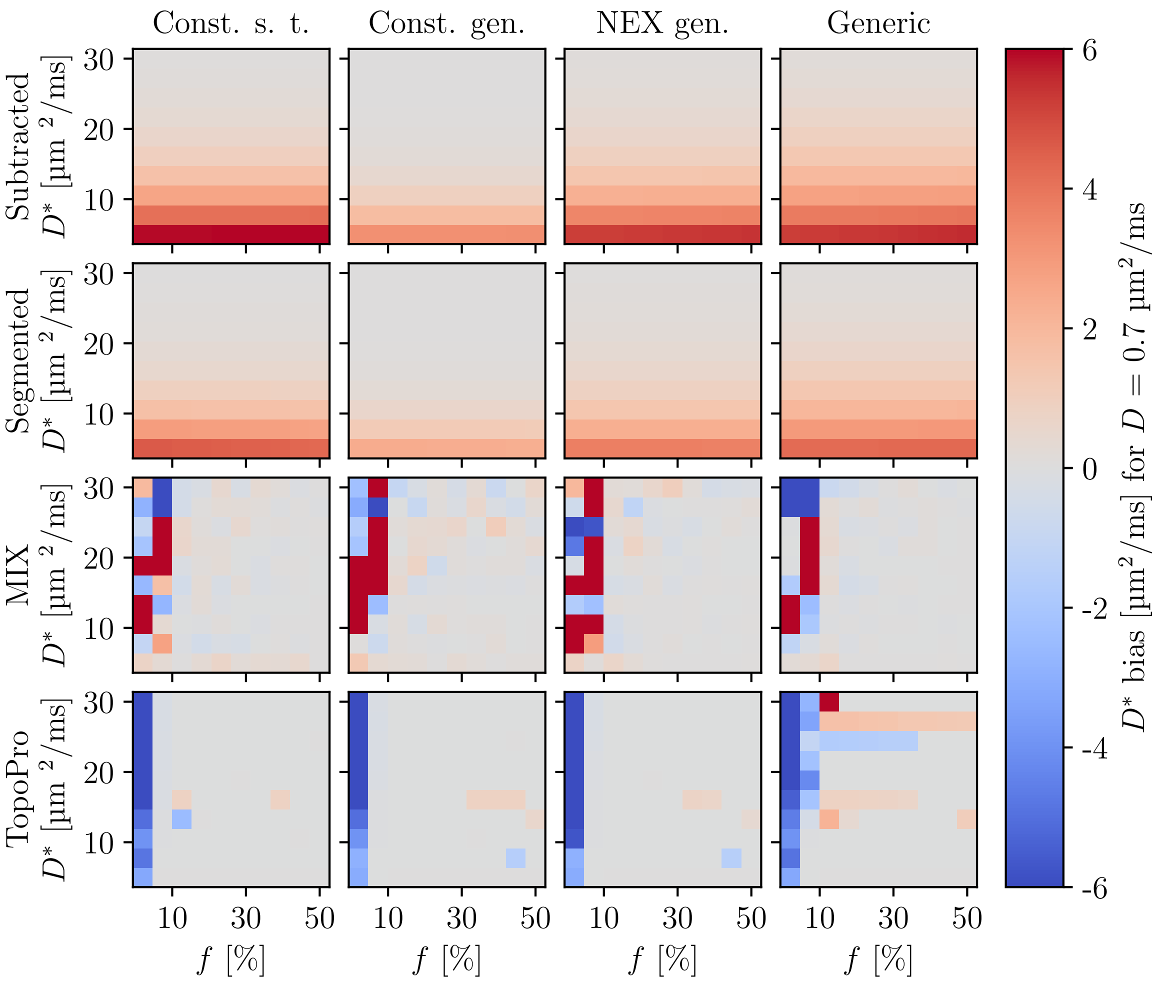

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D* [μm ²/ms]
30
20
10
10
30
50
f [%]
20
13
7
0
-14
-29
-44
f bias [%] for D = 3.0 μm²/ms

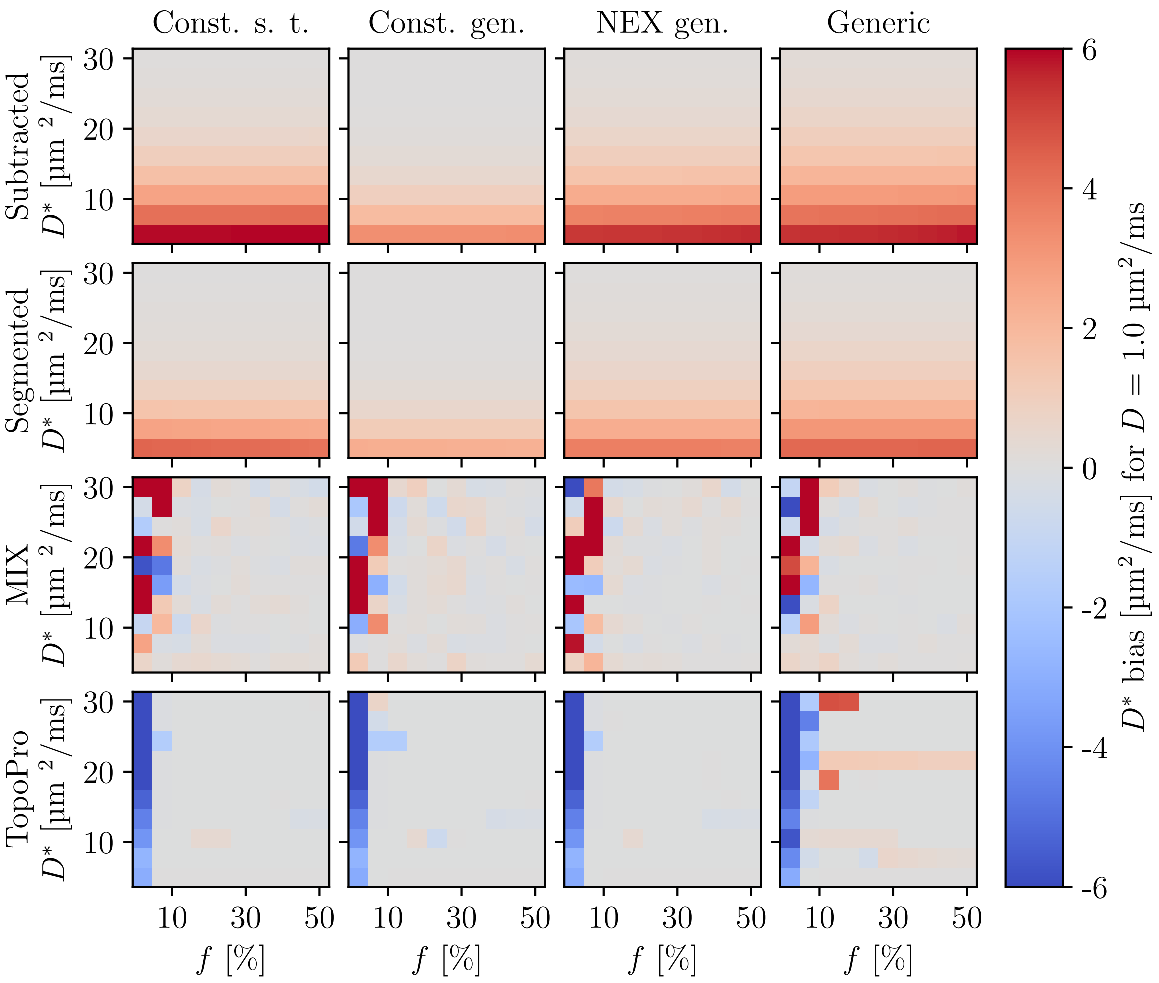
D.2 Bias in D^*

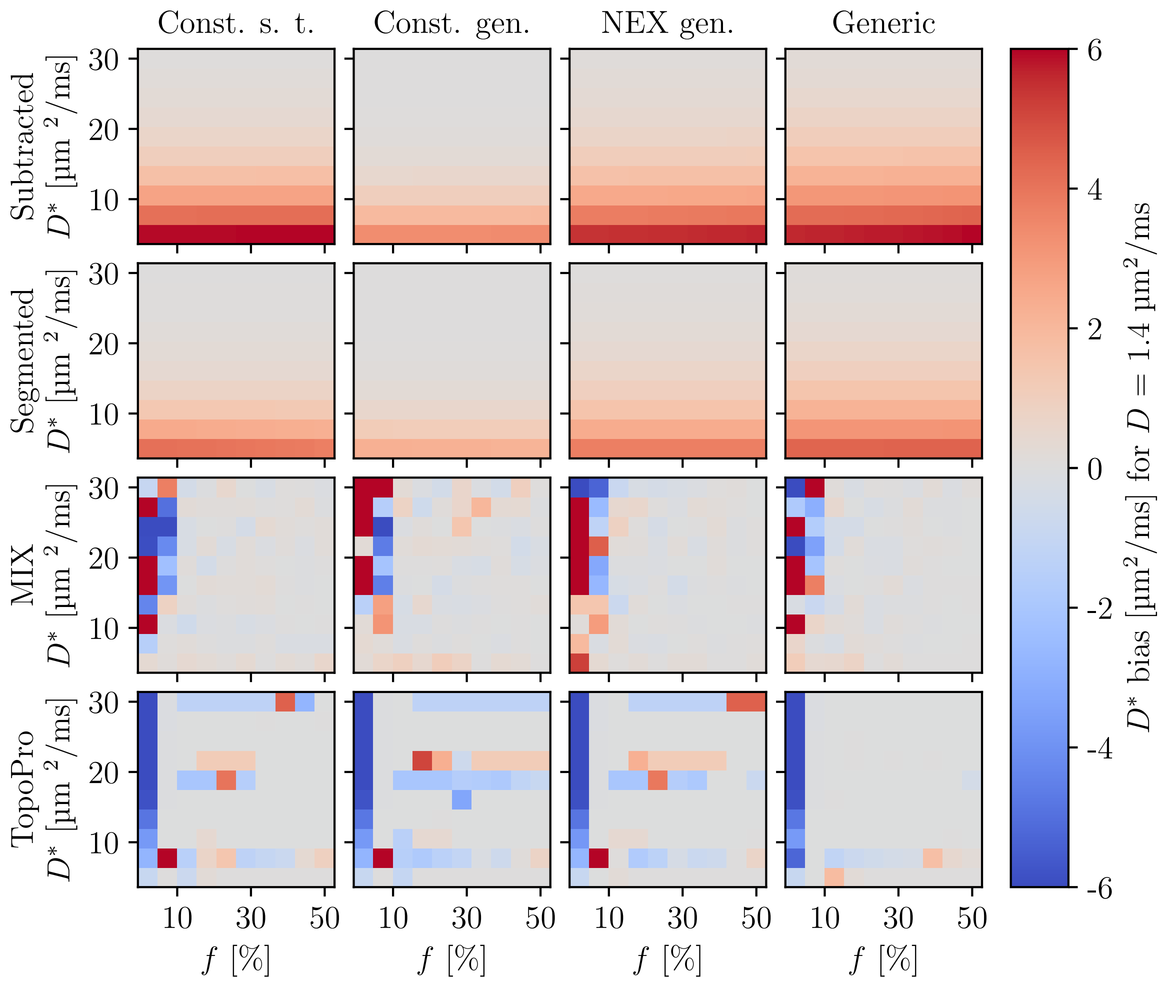
In this section, bias maps for the D^* estimates are presented for all tested protocols, with varying ground truth values of D in the slice direction. Note that the scales for the positive (red) and negative (blue) colours differ from each other.

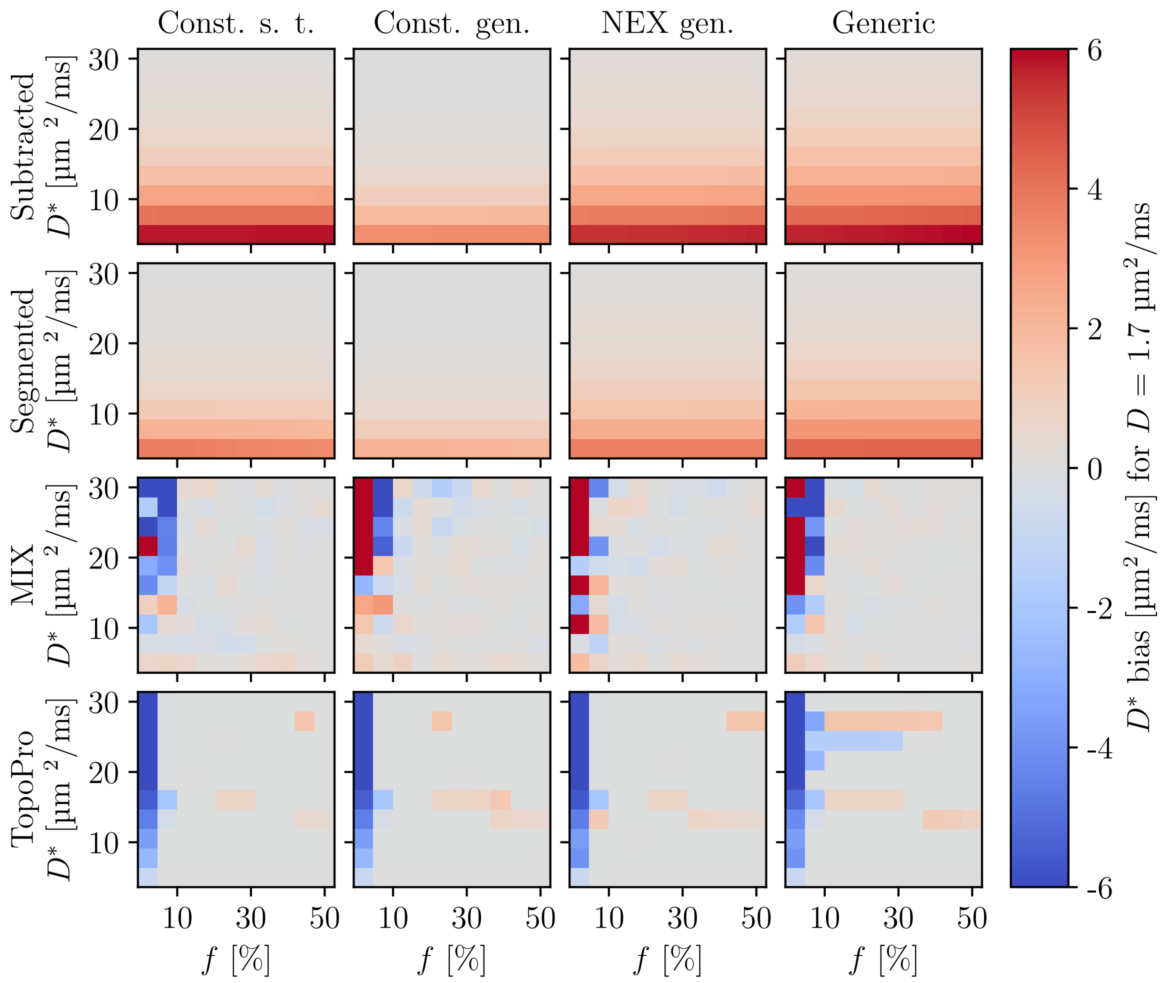

Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
D^* bias [µm^2/ms] for D = 0.1 µm^2/ms
6
4
2
0
-2
-4
-6

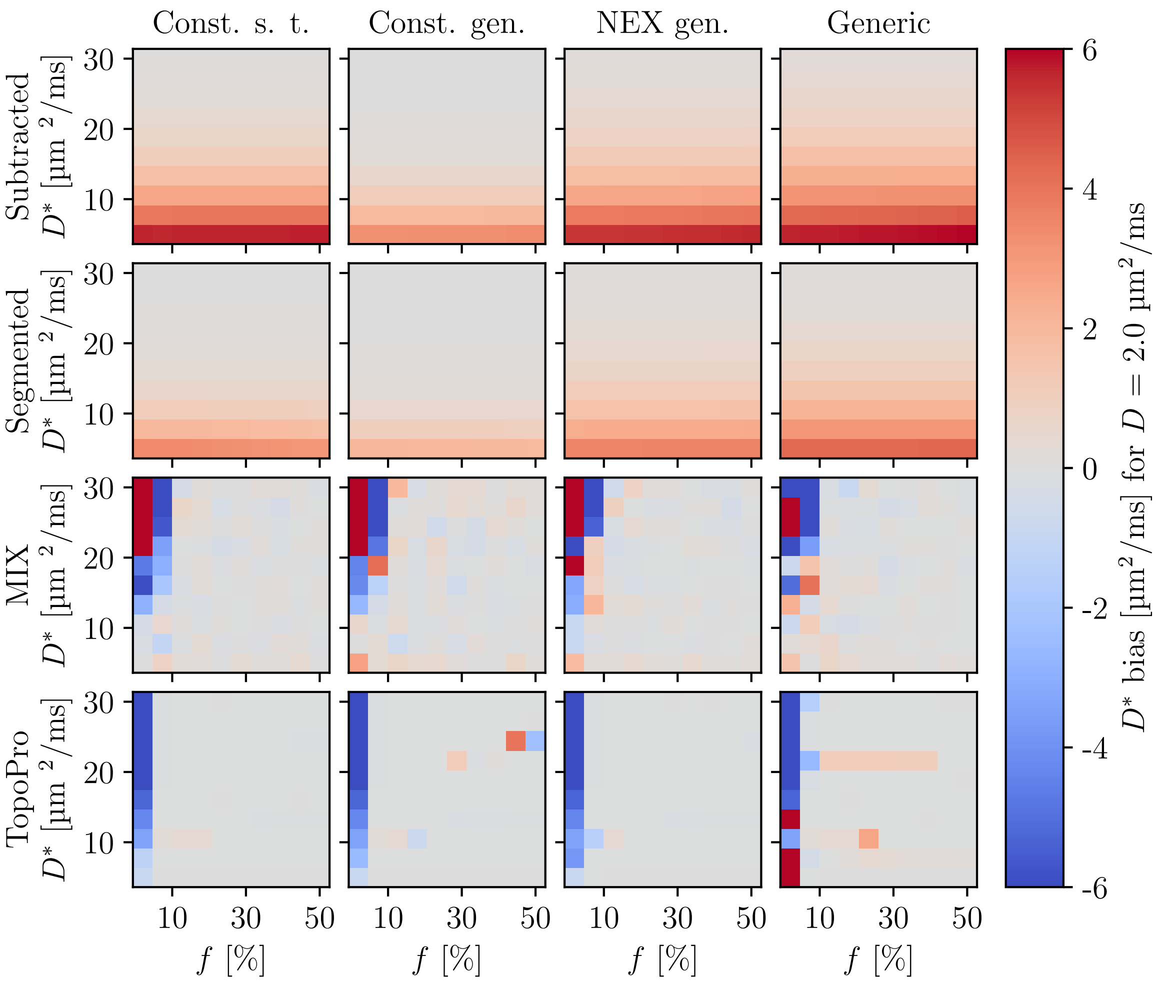

Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
30
20
10
10
30
50
f [%]
6
4
2
0
-2
-4
-6
D^* bias [μm^2/ms] for D = 0.4 μm^2/ms

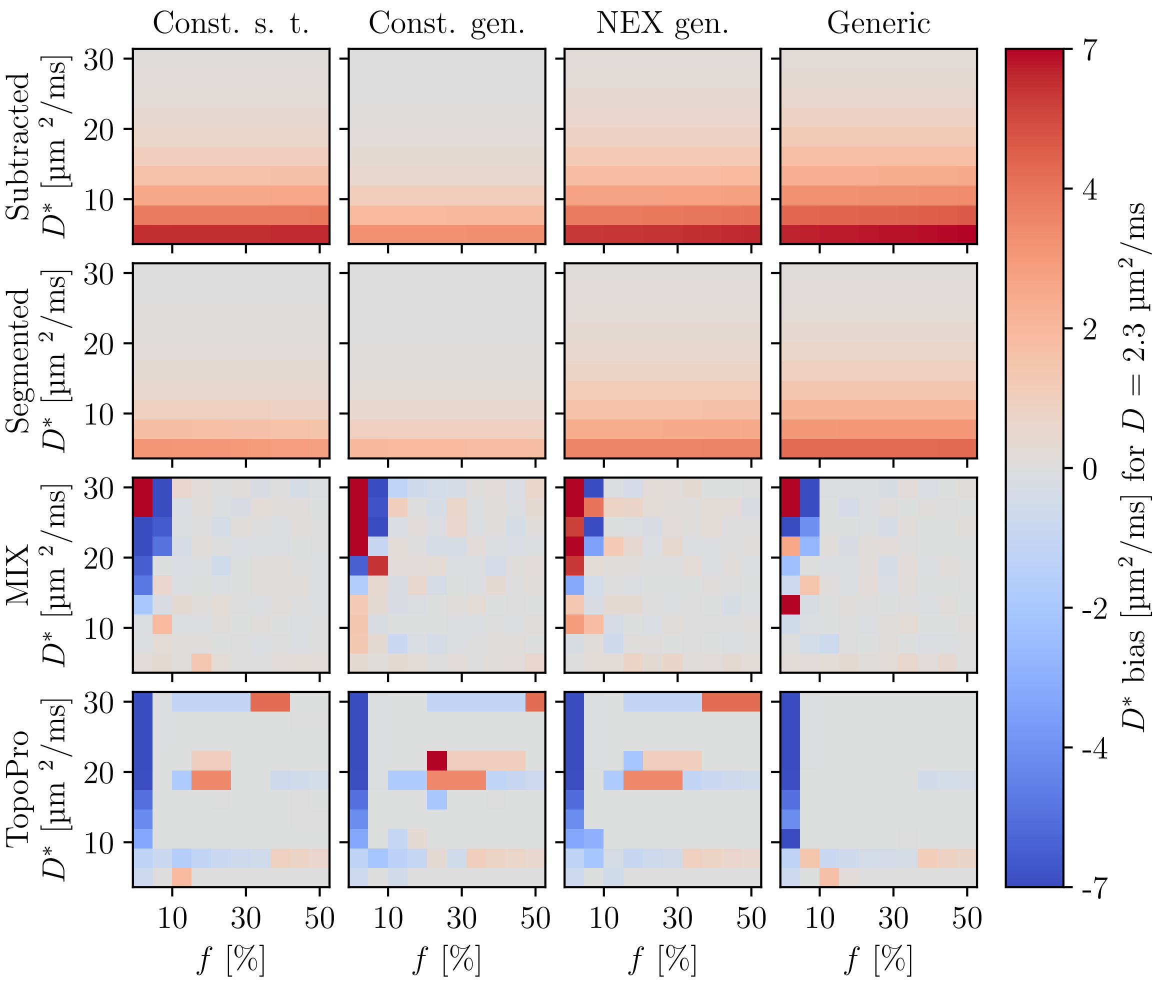

Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
10
20
30
f [%]
10
30
50
D^* bias [µm^2/ms] for D = 0.7 µm^2/ms
6
4
2
0
-2
-4
-6

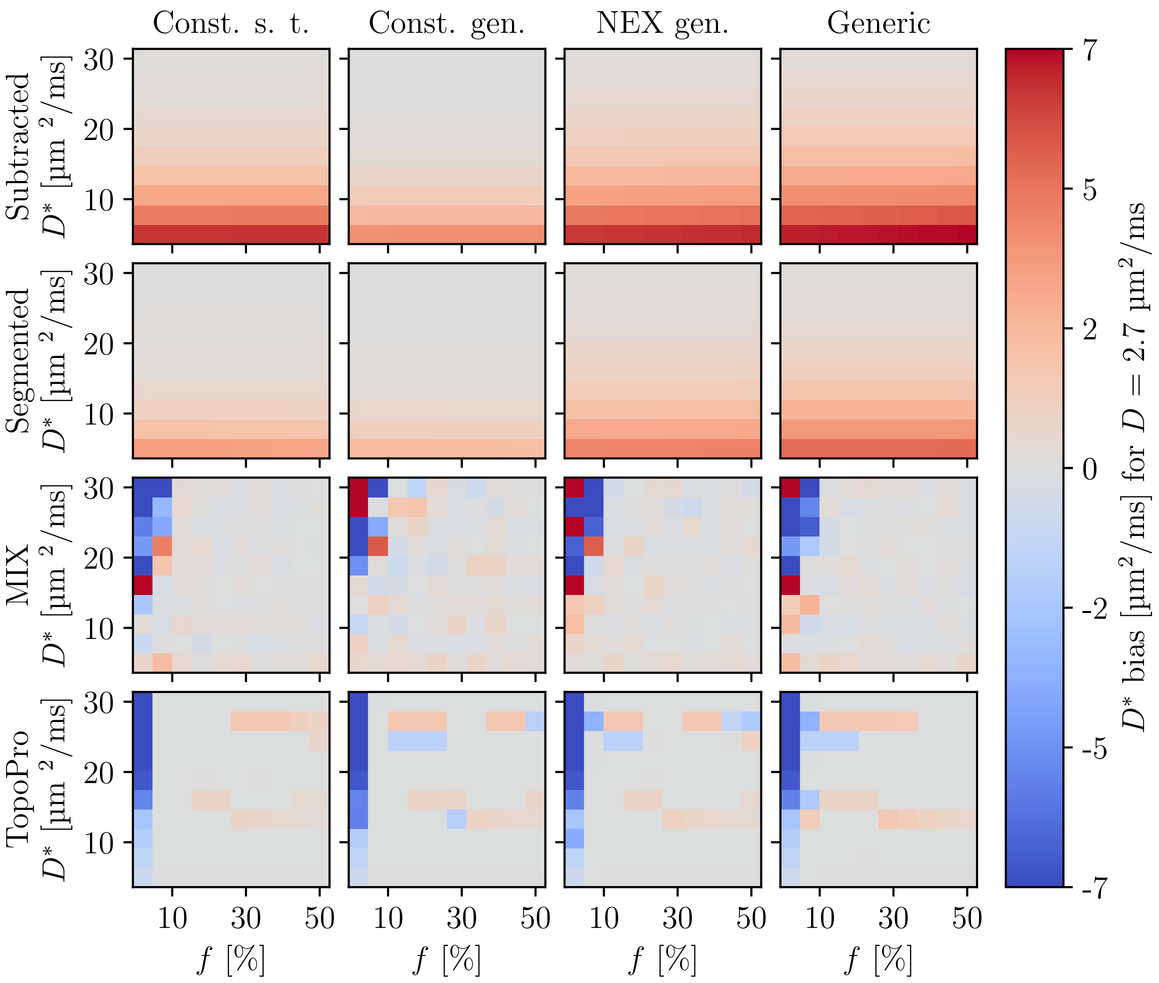

Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
D^* bias [µm^2/ms] for D = 1.0 µm^2/ms
6
4
2
0
-2
-4
-6

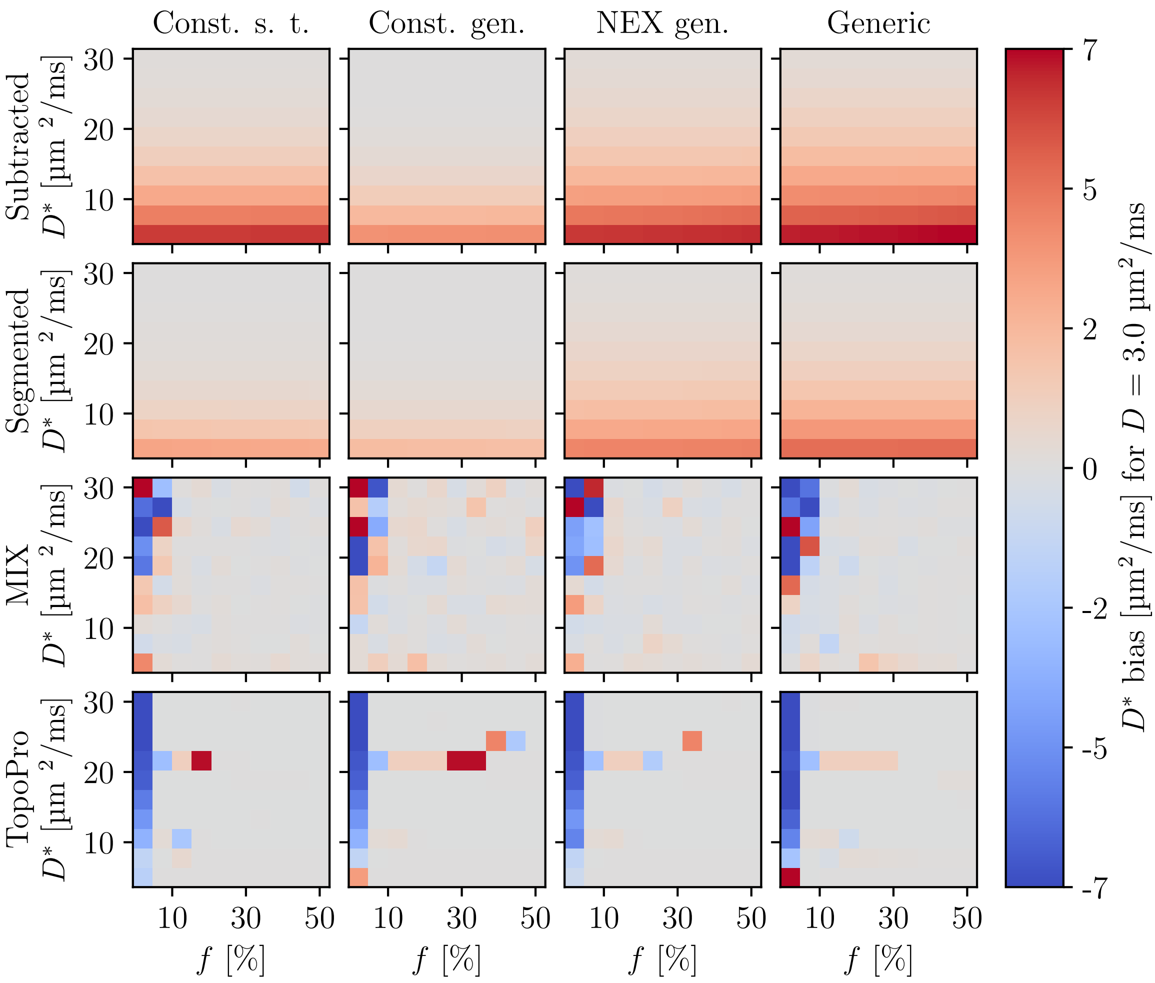

Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
30
20
10
10
30
50
f [%]
6
4
2
0
-2
-4
-6
D^* bias [μm^2/ms] for D = 1.4 μm^2/ms


Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
D^* [µm ^{2}/ms]
Segmented
D^* [µm ^{2}/ms]
MIX
D^* [µm ^{2}/ms]
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
6
4
2
0
-2
-4
-6
D^* bias [µm^2/ms] for D = 1.7 µm^2/ms


Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
30
20
10
10
30
50
f [%]
6
4
2
0
-2
-4
-6
D^* bias [μm^2/ms] for D = 2.0 μm^2/ms

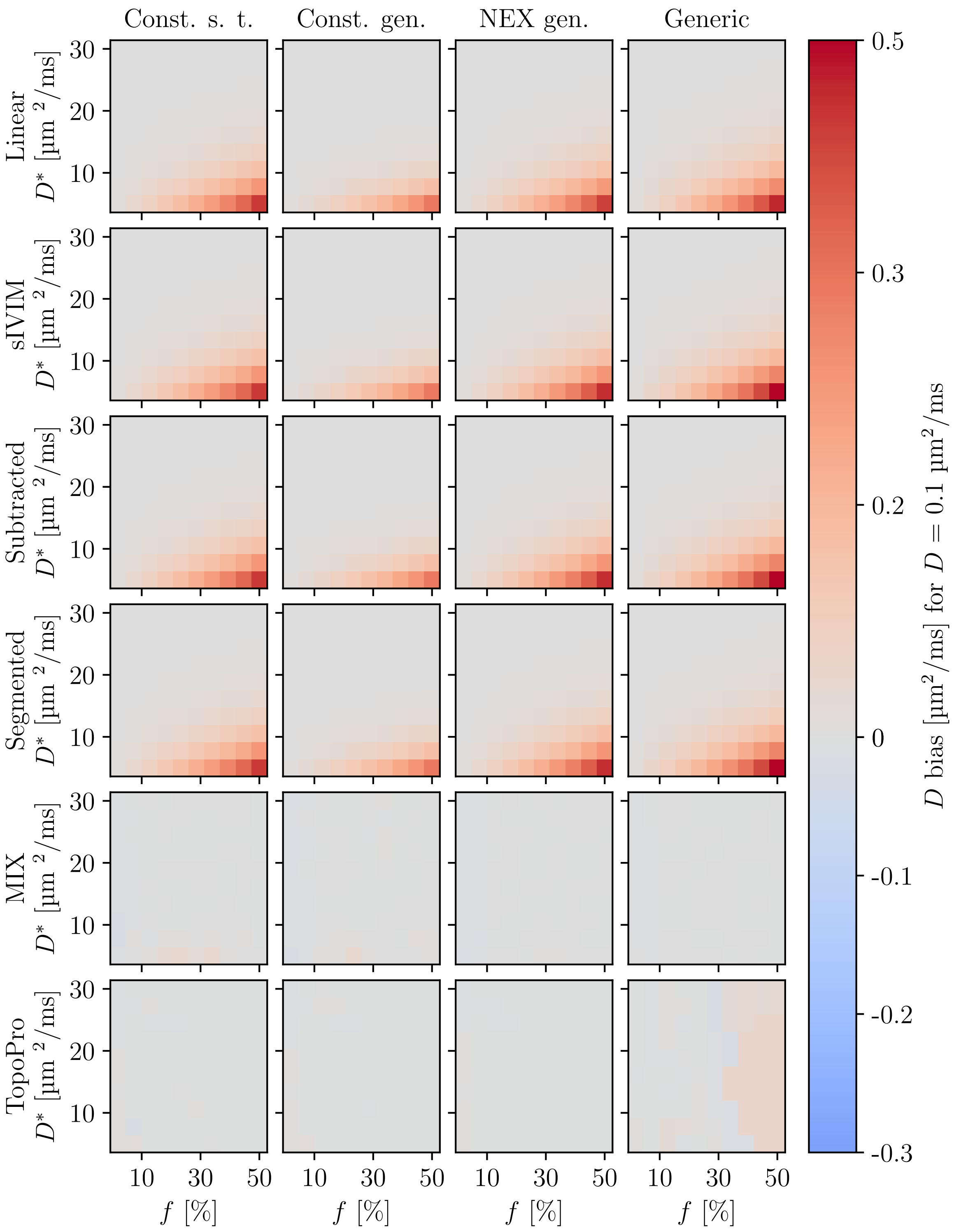

Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
30
20
10
10
30
50
f [%]
D^* bias [μm^2/ms] for D = 2.3 μm^2/ms
7
4
2
0
-2
-4
-7

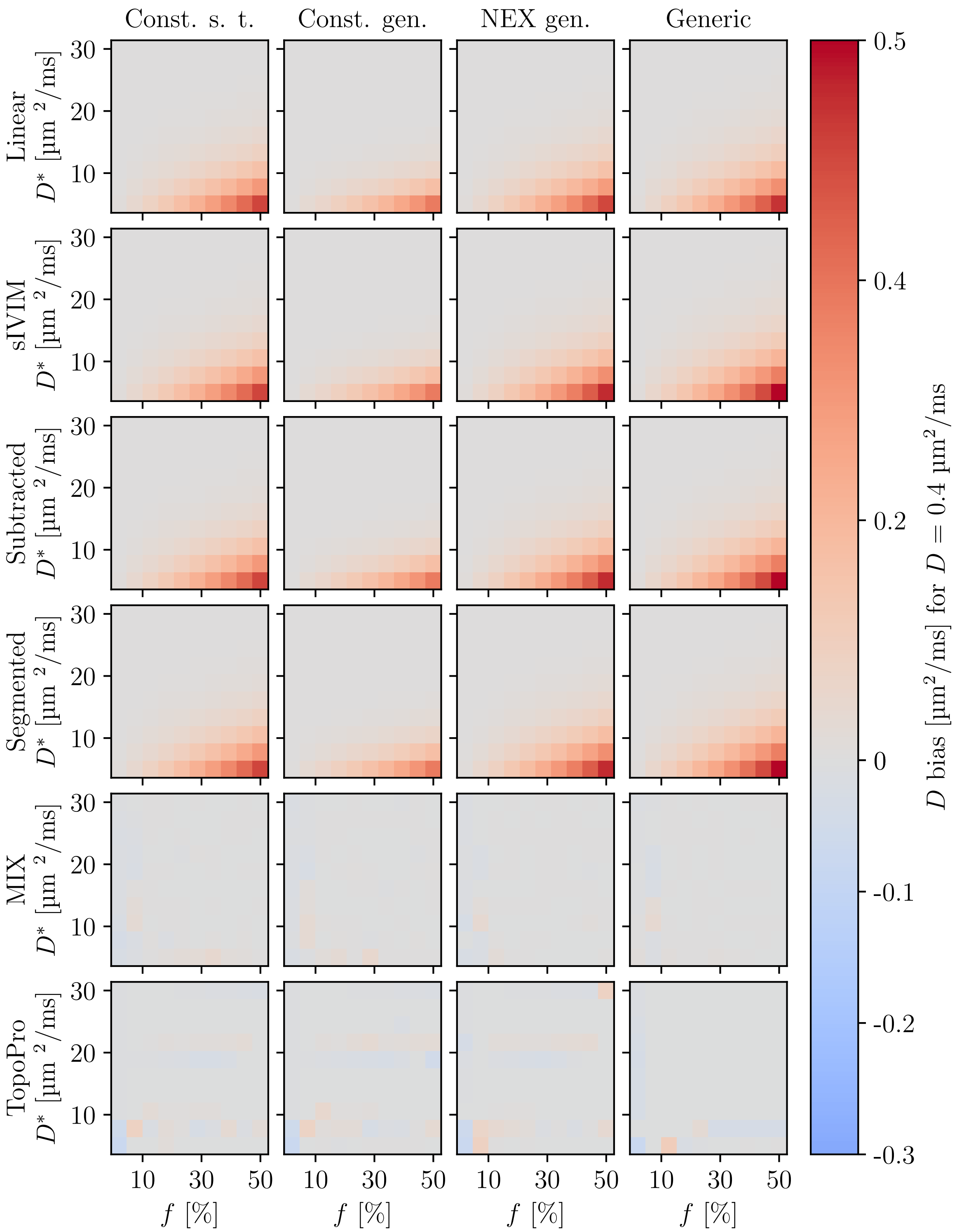

Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
D^* bias [µm^2/ms] for $D = 2.7$ µm^2/ms
7
5
2
0
-2
-5
-7

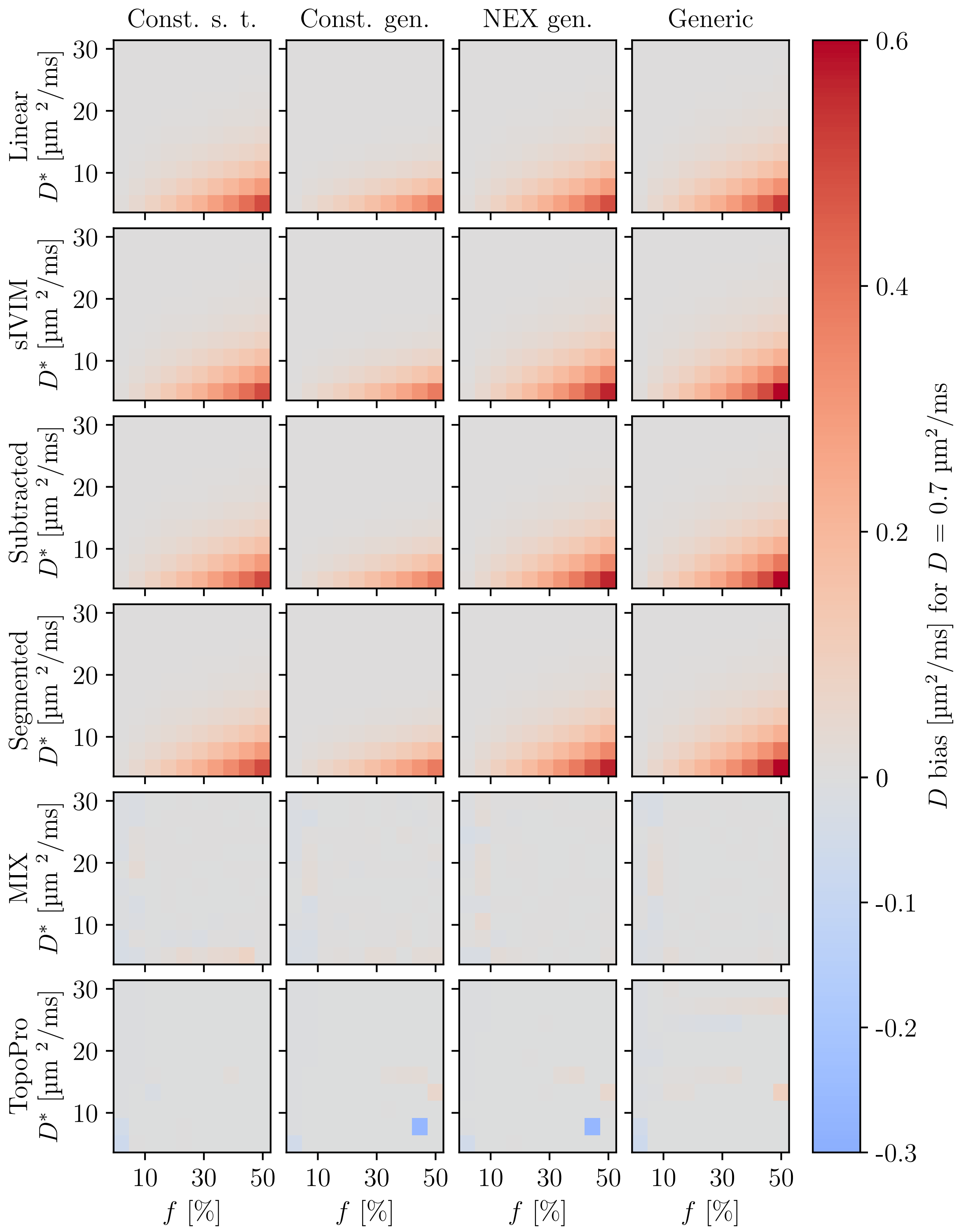

Const. s. t.
Const. gen.
NEX gen.
Generic
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
7
5
2
0
-2
-5
-7
D^* bias [µm^2/ms] for D = 3.0 µm^2/ms

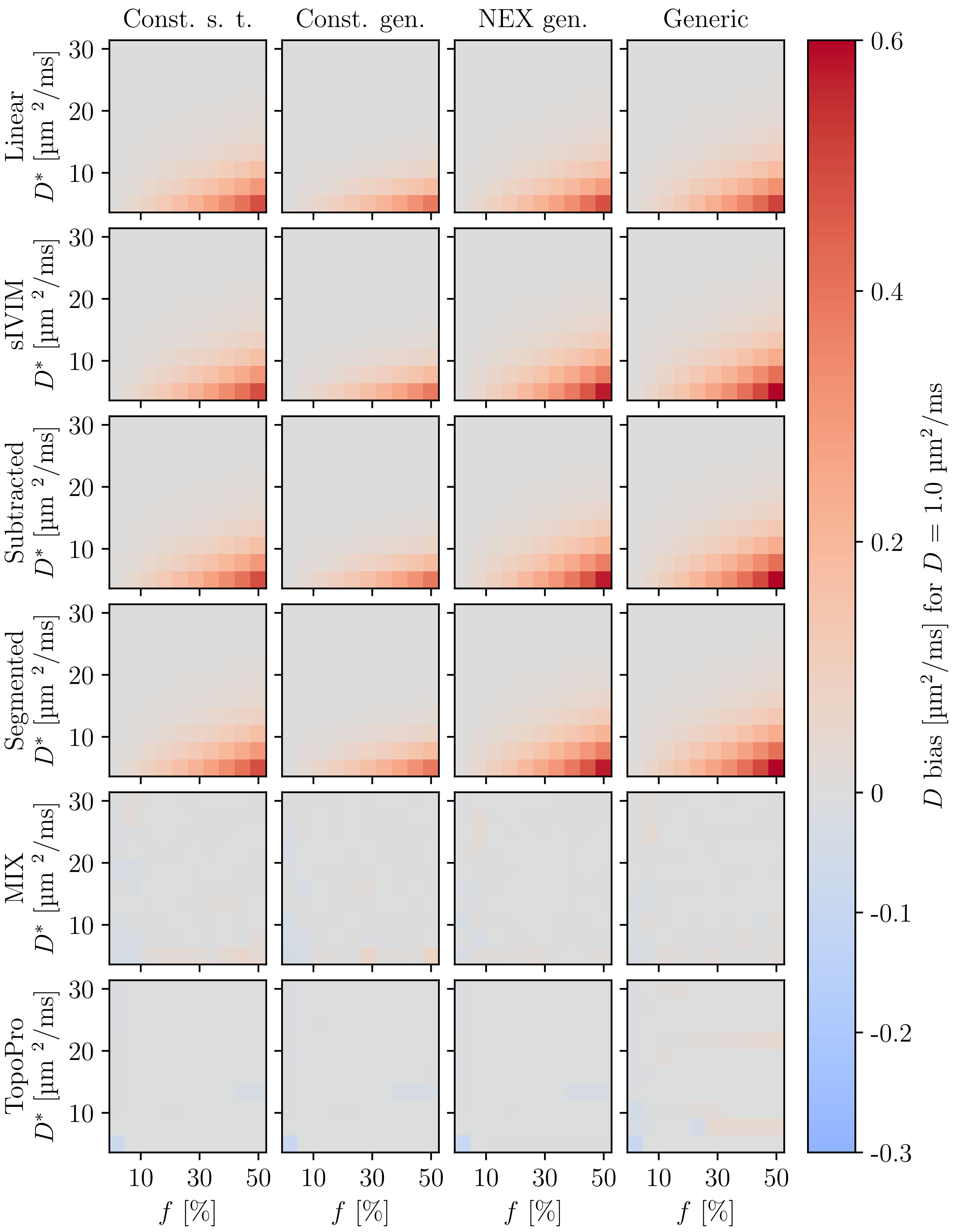
D.3 Bias in D

In this section, bias maps for the D estimates are presented for all tested protocols, with varying ground truth values of D in the slice direction. Note that the scales for the positive (red) and negative (blue) colours differ from each other.

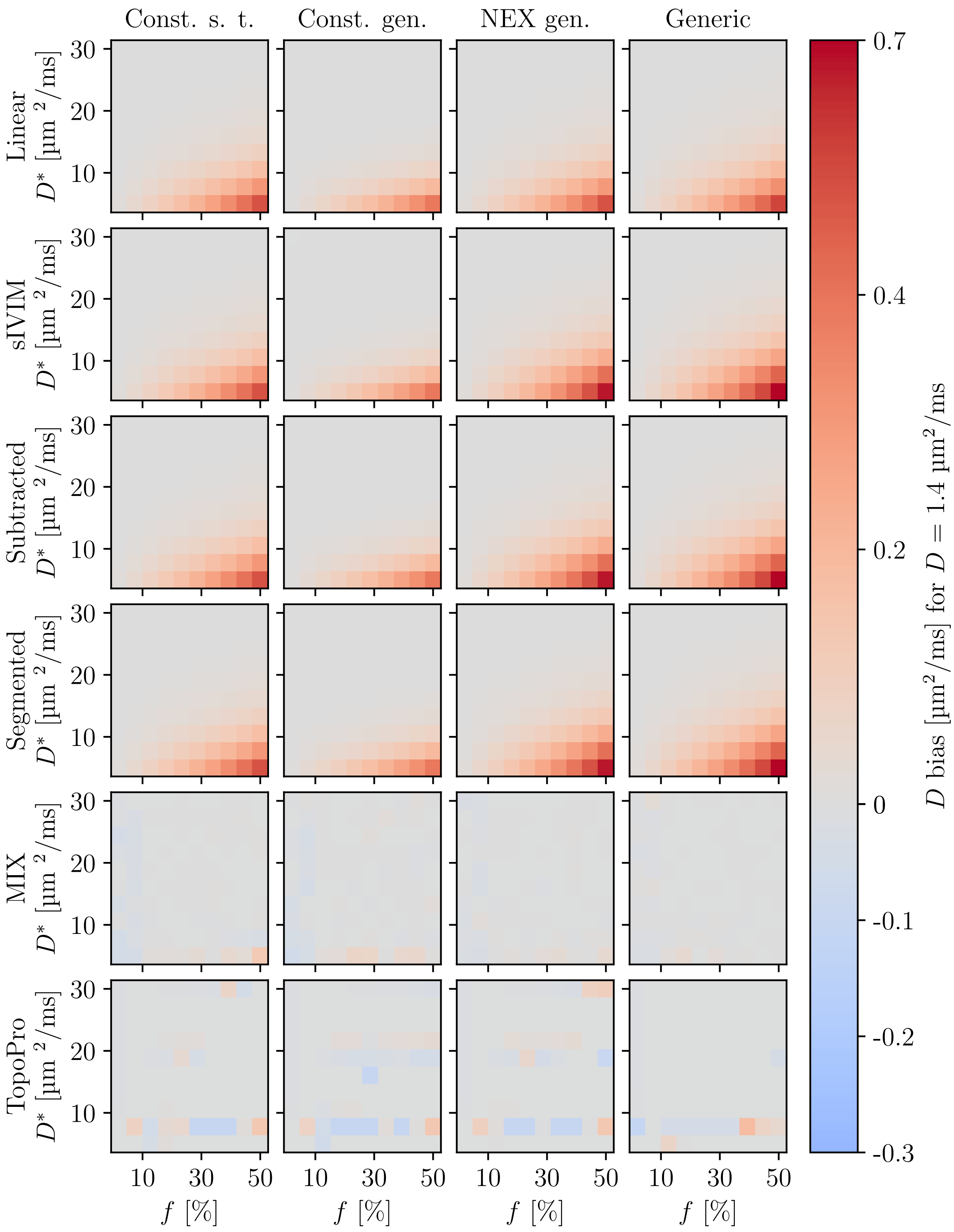

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
D bias [µm^2/ms] for $D = 0.1$ µm^2/ms
0.5
0.3
0.2
0
-0.1
-0.2
-0.3

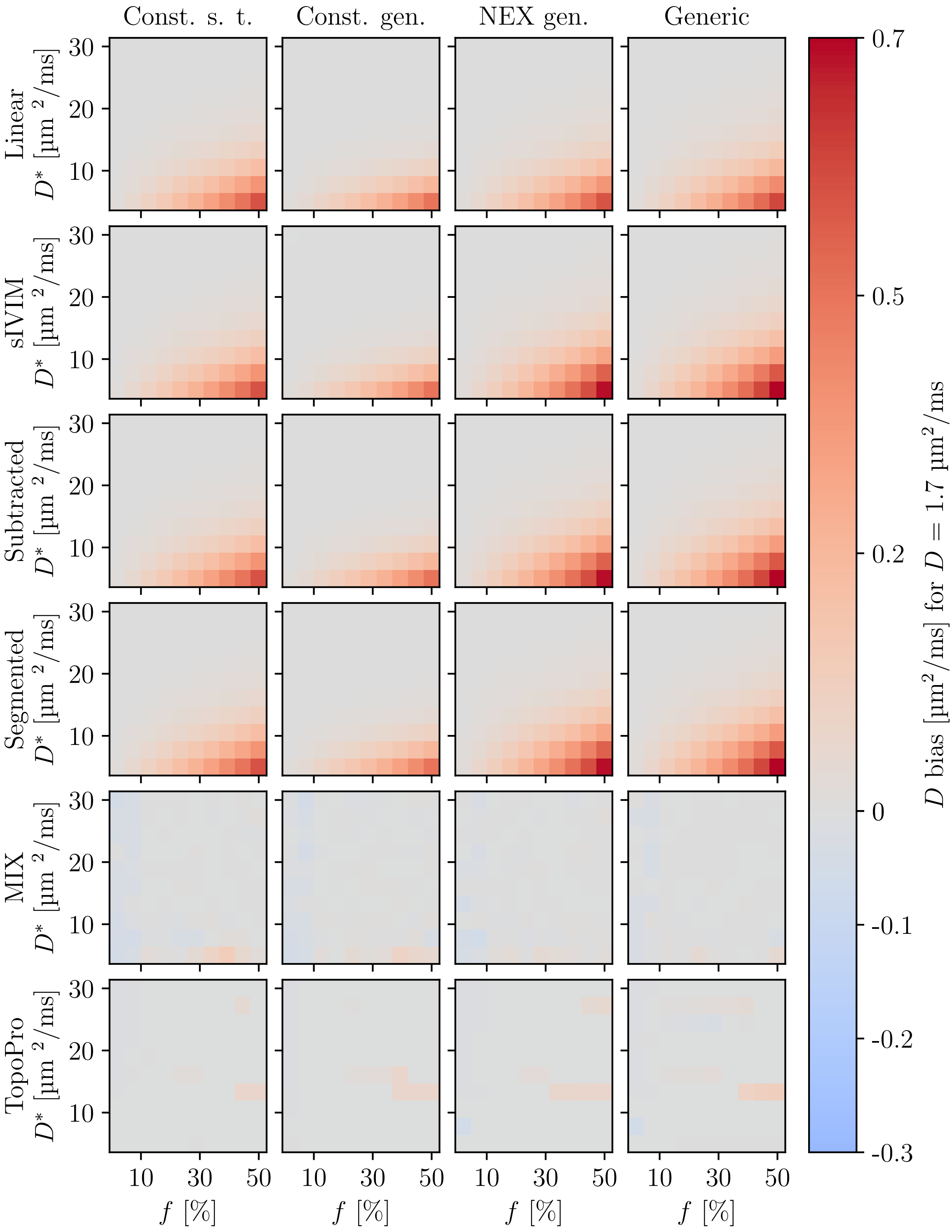

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
0.5
0.4
0.2
0
-0.1
-0.2
-0.3
D bias [µm^2/ms] for $D = 0.4$ µm^2/ms

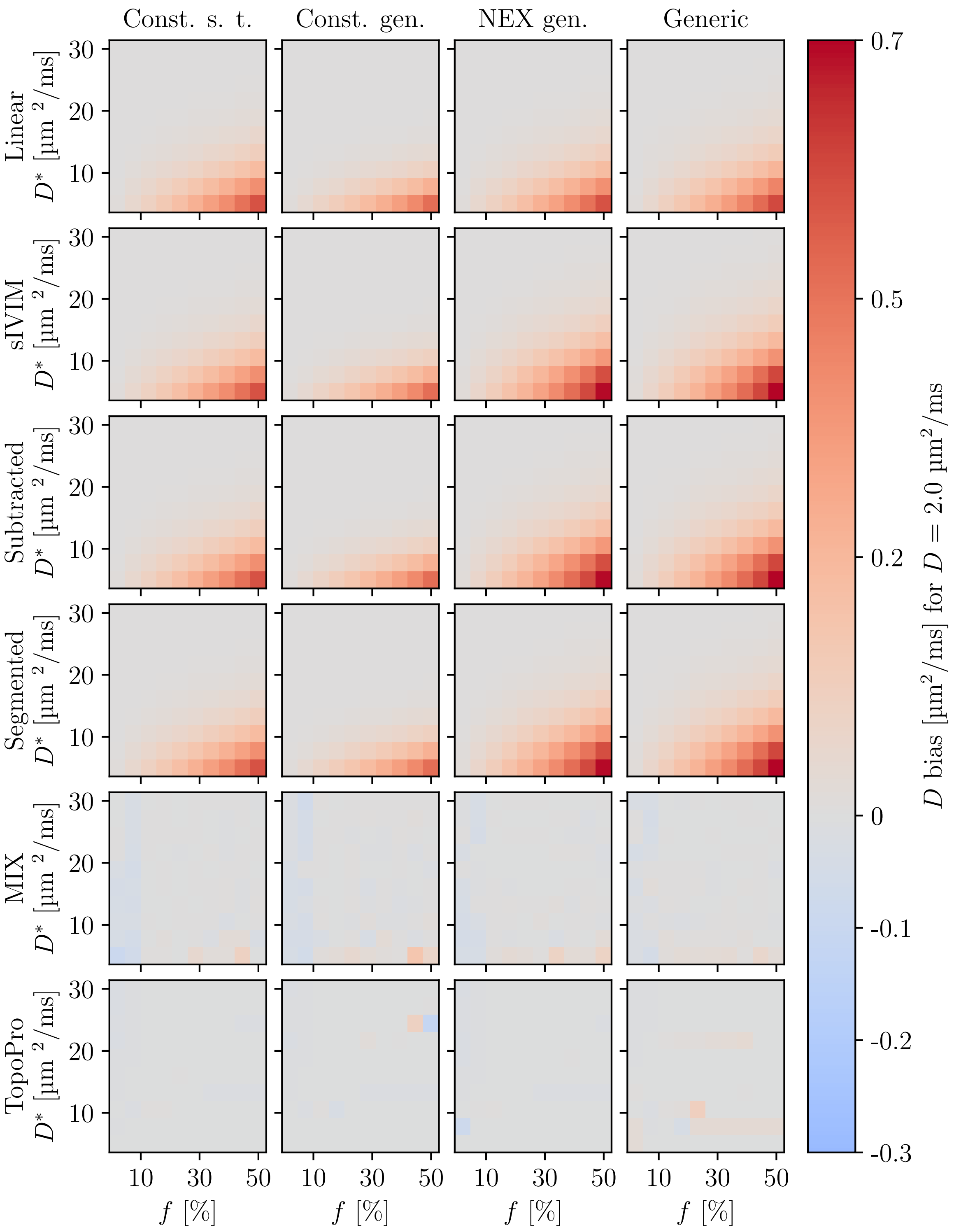

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
0.6
0.4
0.2
0
-0.1
-0.2
-0.3
D bias [µm^2/ms] for D = 0.7 µm^2/ms

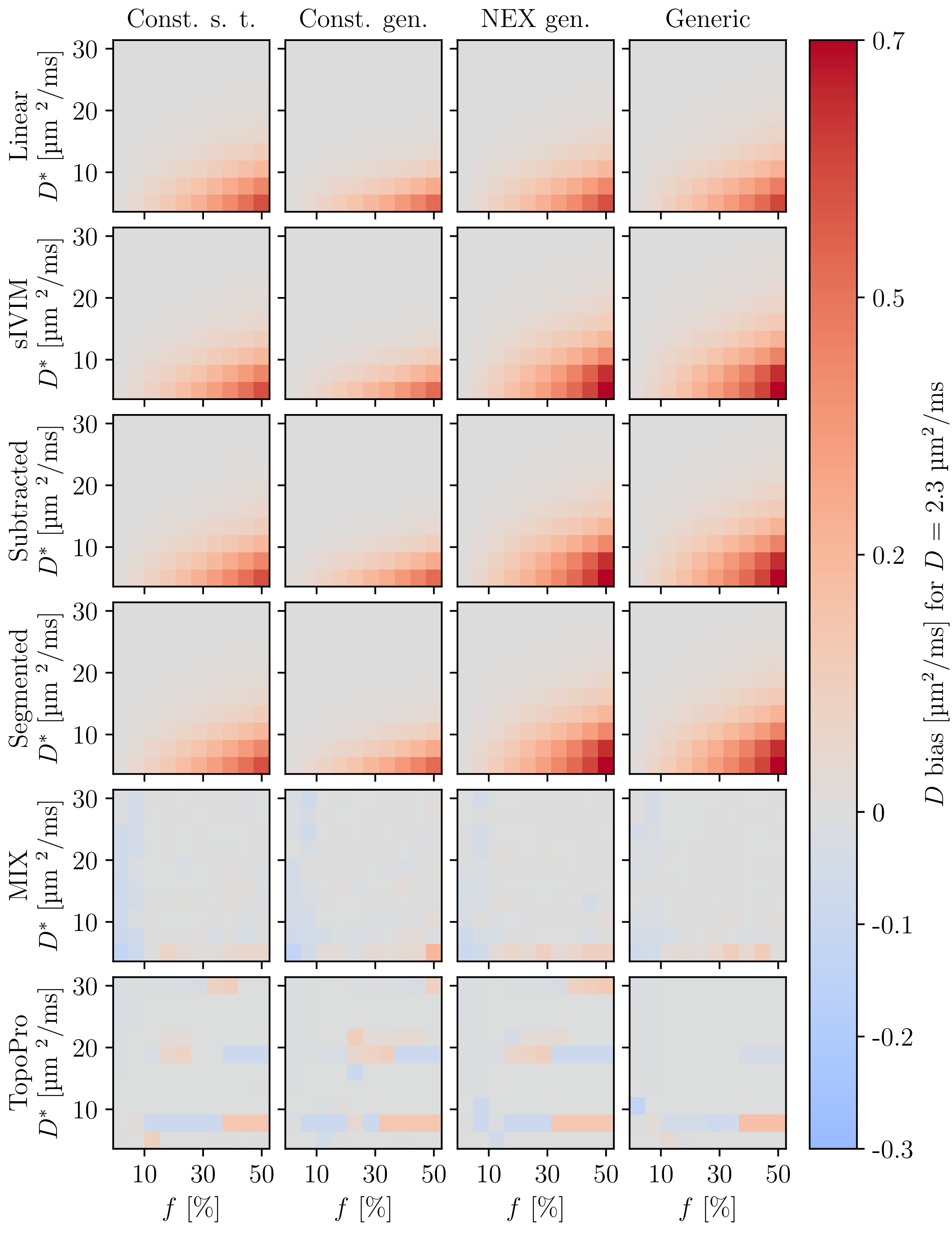


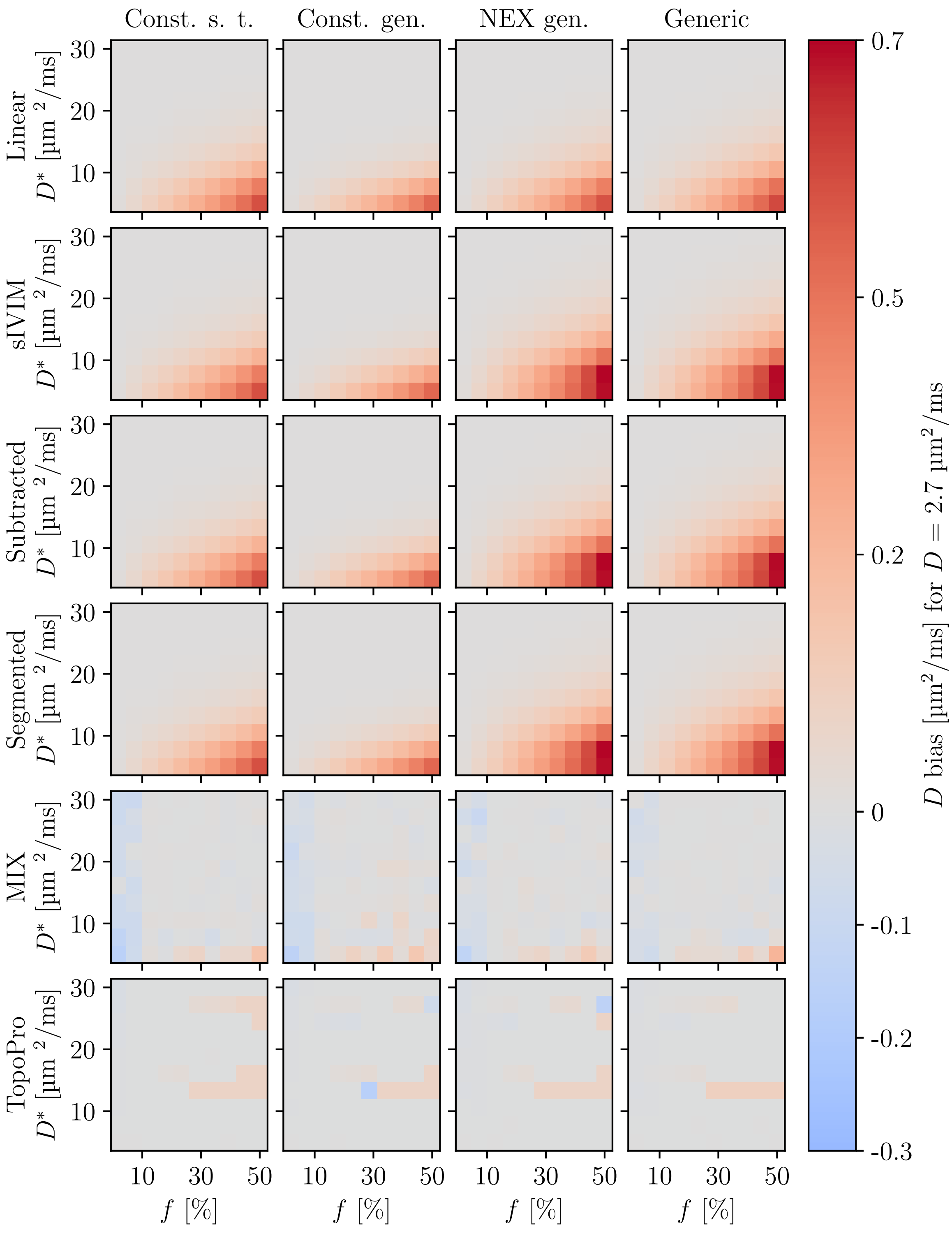
Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
30
20
10
10
30
50
f [%]
D bias [μm^2/ms] for $D = 1.0$ μm^2/ms
0.6
0.4
0.2
0
-0.1
-0.2
-0.3

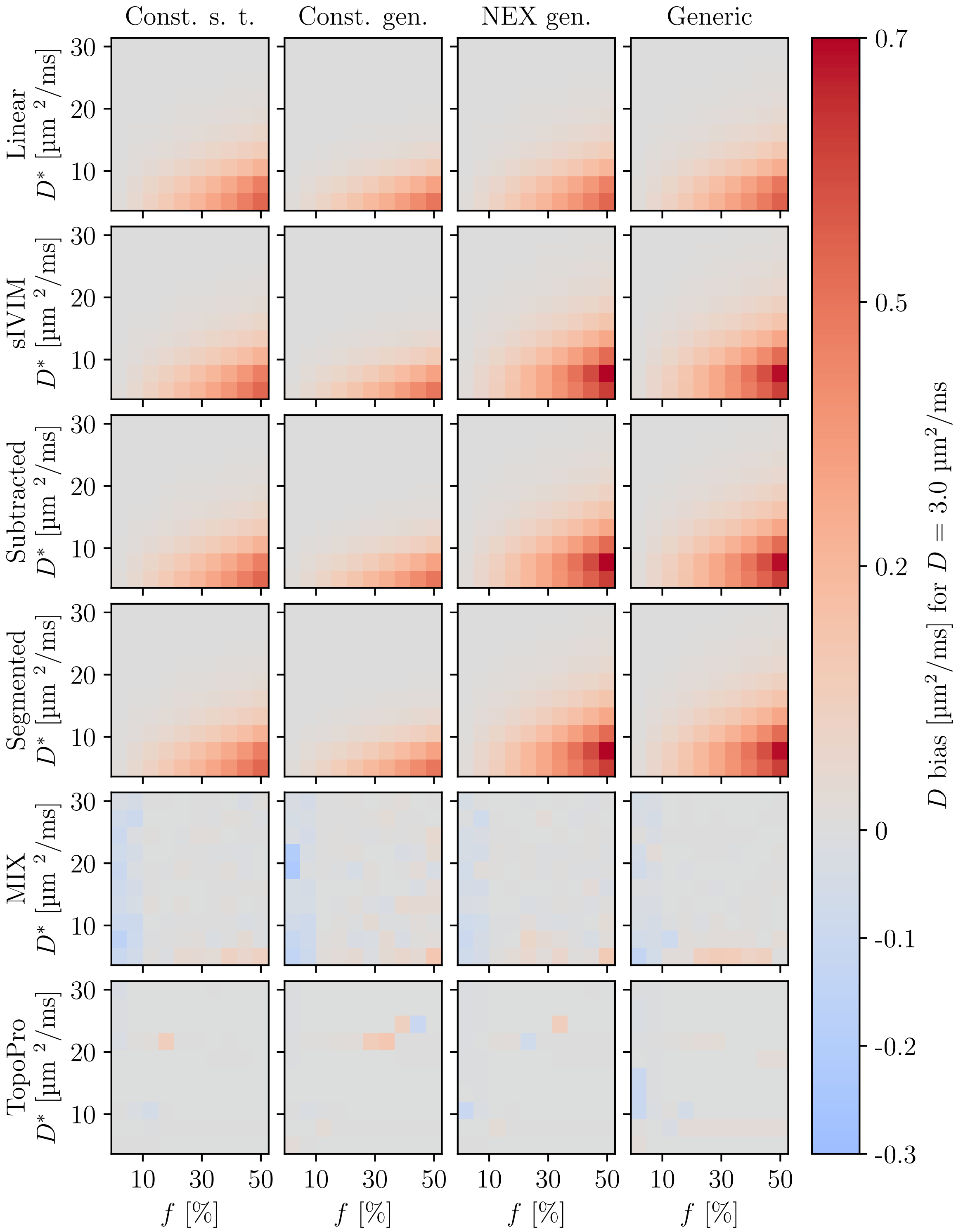

Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [μm ^{2}/ms]
30
20
10
10
30
50
f [%]
D bias [μm^2/ms] for $D = 1.4$ μm^2/ms
0.7
0.4
0.2
0
-0.1
-0.2
-0.3


Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
f [%]
10
30
50
D bias [µm^2/ms] for $D = 1.7$ µm^2/ms
0.7
0.5
0.2
0
-0.1
-0.2
-0.3


Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
10
20
30
10
30
50
f [%]
0.7
0.5
0.2
0
-0.1
-0.2
-0.3
D bias [µm^2/ms] for D = 2.0 µm^2/ms


Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
10
20
30
10
30
50
f [%]
0.7
0.5
0.2
0
-0.1
-0.2
-0.3
D bias [µm^2/ms] for $D = 2.3$ µm^2/ms


Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
10
20
30
f [%]
10
30
50
D bias [µm^2/ms] for $D = 2.7$ µm^2/ms
0.7
0.5
0.2
0
-0.1
-0.2
-0.3


Const. s. t.
Const. gen.
NEX gen.
Generic
Linear
sIVIM
Subtracted
Segmented
MIX
TopoPro
D^* [µm ^{2}/ms]
30
20
10
10
30
50
f [%]
0.7
0.5
0.2
0
-0.1
-0.2
-0.3
D bias [µm^2/ms] for $D = 3.0$ µm^2/ms