Using Transfer Entropy for Analyzing Biological Networks

Master Thesis

Presented to the Faculty of Biosciences of the Ruprecht-Karls-Universität Heidelberg

Name:	Sirac Baz
Matriculation Nr.:	2859247
Supervisors:	Prof. Dr. Ursula Kummer
	Dr. Jürgen Pahle
Date:	07.05.2019

This Thesis was written at the Institute of Bioquant at the Ruprecht-Karls-Universität Heidelberg in the period from 1st of August 2018 to 7th of May 2019 under the supervision of Dr. Jürgen Pahle.

1st Examiner:	Prof. Dr. Ursula Kummer	Institute:	COS / BioQuant
2nd Examiner:	Dr. Jürgen Pahle	Institute:	COS / BioQuant

I herewith declare that I wrote this Masters Thesis independently, under supervision, and that I used no other sources and aids than those indicated throughout the thesis.

_____ _____

Date Signature

Acknowledgment

I want to thank Dr. Jürgen Pahle for motivating me to face this topic, for our discussions and his advice. I also want to thank Prof. Dr. Ursulla Kummer that she accepted to act as a supervisor for my thesis.

My thanks go to Arne Schoch as we could share the experience of writing our scientific stuff at the institute, instead of sitting alone there and being bored.

My special thanks go to Ludwig Richter who showed me how this latex stuff works. This work would `look so awful` if it were not for him; a really supportive and good friend.

I must thank Lennart Srelha that he supported and motivated me to go on, as he also worked in parallel to finish his final exams. He makes good caricatures about the subjective misery of scientific work (just look at the whiteboard).

Many thanks go to all the lovely people for being just there as Lukas, Nila, Laura, Luggi, Lennart, Anielle and Arne. They were and are really supportive to me, as scientific work can be very isolating, and good talk is not cheap.

Finally, I want to thank Anielle for being on my side.

Abstract

One of the most important goals of systems biology is to understand cell signaling on a systemic level. As signaling pathways are crucial for cellular communication, decision-making and regulation of biological functions, quantitative and computational methods are used to shed light on these networks. In this context, the information flow in biological structures can be studied using information-theoretic measures. Information theory provides tools to understand information storage, encoding, processing and transfer without being dependent on prior knowledge about the system. Such model-free approaches quantify the statistical dependencies between processes and are used upon synthetic as well as real-world data. One application of information theory is network analysis in order to gain insights into the structural properties of biological systems. Numerous research demonstrates the widespread use of information theory upon biological data to infer causal relationships. Especially transfer entropy stands out as a beneficial tool to detect causal relations in a non-parametric way.

The novel approach of this thesis is the application of conditioned transfer entropy to study biological signaling networks. The purpose is to assess the possibilities and caveats of inferring the causal relations in signaling system by using transfer entropy. In particular, conditioned transfer entropy allows to eliminate non-causal information transfer, hence it is useful to quantify direct correlations and enables causal inference.

The quantification of information transfer requires to estimate the probability distributions of given variables. The validity of these estimations were assessed by using surrogate data. The MAPK cascade was used as a test case for a transfer entropy analysis of connectivity, as its structure is a recurrent building block of many cell signaling pathways. Additionally, simple and generic models were created to assesses the reliability and robustness of transfer entropy estimations. Different properties of the data and important estimation parameters, such as the data length, sampling frequency and embedding dimensions, were scanned to evaluate the impact on information transfer.

The outcome demonstrates the difficulty to eliminate non-causal correlations and infer the underlying structures. No practical guarantees can be given that non-causal information transfer is sufficiently blocked by conditioned transfer entropy. However, different criteria to perform a meaningful information transfer analysis in the context of causal inference could be derived. The behavior of transfer entropy upon varying embedding dimensions can be used to identify appropriate sampling frequencies for a

information transfer analysis of connectivity. Such sampling frequencies can also hint to the time scales which govern the causal and non-causal correlations between processes. The surrogate estimations can be used as indicators of the estimator bias and help to identify the parameter sets which decrease such bias. In this context, the application of an improved estimator exhibiting less bias is recommended, especially in the case of highly dependent variables. This is particularly important if one wants to analyze biological networks in which the processes are usually highly interdependent. Moreover, the data length must be only sufficiently high to exhibit the qualitative differences between various source-target constellations. Full convergence of information transfer by very high data lengths is not necessary for a qualitative inference of connectivity.
Finally, I discuss the possibilities to combine the criteria for transfer entropy analysis developed here with criteria evolved in the literature. As non-causal information transfer remains a part of transfer entropy analysis, I also argue how a redirection of transfer entropy can be fruitful to quantify the emergent computational properties of a system.

Contents

1. Introduction

Cell signaling is one of the most important features of biological systems which guarantee the ability of living organisms to process information coming from their environment, to establish a molecular communication between cellular components and inside of the cell, to regulate biological functions and to realize decision making processes. We can also regard biochemical signaling systems as complex relay stations which store, process and transfer information across the cell and organism and connect different (sub-)systems with each other. From that illustration it becomes clear why it is so attractive to use a theory which is capable of describing the notion of information and related concepts precisely, as it is accomplished by information theory [1]. Especially in systems biology, where the objective is to gain insights into the system-level properties of biological networks [2], the application of computational methods to decipher the dynamics and functions of signaling pathways requires solid theoretical foundations. In this context, information theory is an exact and flexible framework which provides various tools to analyze information flow with statistical means. In the past years, more and more information-theoretic measures were integrated in research in order to understand biological networks which store, process, encode and transfer information [3]. Such techniques are applied on synthetic as well as real-world data [4]. In particular, cellular signaling is subject to a wide variety of techniques based on information theory which are used to understand different aspects such as decision making [5], recurrent motifs [6], cross-talk [7] and topology [8]. An extensive overview about different mechanism in signal transduction can be found in [9, 10].

One prominent biological application of information theory is network analysis where the inference of the underlying topology of a system is the goal [11]. Particularly in the fields of neuroscience [12] and genetics [13], information-theoretic approaches are employed to calculate statistical dependencies between different (sub-)systems using experimental data. The benefit of these methods is that they are model-free, hence no previous knowledge about the system has to be included in the analysis [14]. In this regard, my thesis contributes to the approach of inferring the underlying structure of signaling networks by

using information-theoretic measures. Network analysis upon signaling pathways were performed in the past, but these were either simulation based [15], or used Bayesian network analysis [16]. My novel approach is to apply conditioned transfer entropy (TE) to elucidate the causal structures of biological signaling. TE is an information-theoretic measure which quantifies the directed information transfer between two processes [17]. Signaling pathways were analyzed via TE before, but in order to decode the impact of changing dynamics upon information transfer [18], not for causal inference. As TE is a dynamic and asymmetric measure, it is an information-theoretic quantity which converges most to the principle of causality. Moreover, conditioned TE can be used to block non-causal information transfer, thereby enabling the inference to direct and causal relations. Hence, the possibilities and caveats of applying TE to gain meaningful insights into the underlying structure of biological networks are evaluated. For this purpose, the Mitogen Activated Protein Kinase (MAPK) pathway serves as a test system, as it is a biologically relevant, evolutionary conserved and well described structure which contains recurrent motifs of feedback loops, phosphorylation cycles and phosphorylation cascades [19, 20]. Concurrently, the MAPK cascade is a simple and small network which renders it an appropriate model to evaluate TE. In this context, an information transfer analysis is performed by applying different conditions upon simulated MAPK data. These conditions concern different properties of the data and parameters of the TE estimation, namely the data length, sampling frequency and embedding dimensions. As TE measures the statistical dependencies between variables to quantify information transfer, probability distributions have to be numerically approximated using estimators. The TE estimations are further validated using a surrogate data method, and surrogate and original estimates are compared. The simulations of the MAPK models are realized via stochastic methods, as stochastic simulation correctly reflects random fluctuations of a system which is also an important prerequisite for TE estimation. The outcome sheds light on the meaningful use of TE for analyzing the connectivity of biological networks. Moreover, criteria for a TE analysis of functional connectivity are derived due to the results of the estimations. Especially the restrictions which come with the use of non-parametric, purely statistical approaches for causal inference, such as TE, have to be taken into account. Against this background, possibilities of redirecting the application of TE are discussed and an outlook for future research is given.

1.1. Information Theory

1.1.1. General Notions on Information Theory

1948, Shannon introduced a general theory of communication to approach the problem of reliable information transfer on a abstract level, given that the transmitting channel is disturbing the signal and the information transfer is independent of the signal to be transmitted [1]. Basically, the theory assumes that a source of information is encoded using a transmitter and delivered to the encoding receiver via a noisy channel [ref:fig].

The most basic quantity in information theory is the Shannon entropy, which is defined as the average number of bits needed to encode independent draws of the discrete distribution X with a probability distribution $p(x_i)$, where x_i indicates all possible states of X over i:

$$H_X = -\sum_i p(x_i) \log_2 p(x_i)$$

The base of the logarithm determines the units of information which typically are bits, nats or hartleys for base 2, the natural logarithm or base 10 respectively. Henceforth, I will use the base 2 logarithm and the respective unit bits.

As information theory is agnostic concerning the transmitted content and approaches the problem of communication from a statistical perspective, it enables the possibility to use it as an analytical tool on a wide range of data. This has been applied in various fields like finance [21, 22], physics [23–25], chemistry [26], computer science [27], medicine [28], meteorology [29], statistics [30], electronics [31] and governance [32] etc., just to name a few.

In biology, the interest for this theory is due to information-theoretic measures which allow to analyze the relation between different processes in respect to their information content [33, 34], statistical dependence (beyond correlation analysis) [35, 36], their decoding and encoding capacity [5, 37] and their structural properties [38], especially concerning the inference of networks [11]. Particularly in two biological fields, gene networks [4, 39, 40] and neuroscience [41–43], information theory was and is received with great interest and found its way into the research of these information processing and transferring structures.

Interestingly, the topic of reliable communication despite of noise was approached by the information-theoretic (Shannon and McCarthy 1956 [44]) and neuroscientific community (Winograd, Goldstine, and Cowan 1963 [45]) about the same time, which shows us how fundamental the question of information transfer is for the field of biology. In this context, different information-theoretic measures, such as mutual information (MI) [46] and channel capacity (CC) [47], have been applied to quantify aspects of information flow.

1.1.2. Transfer Entropy

Schreiber introduced transfer entropy (TE) which is a dynamic and non-symmetric measure to determine the directed information transfer between two processes. In order to analyze their coupling, we can regard the relationship in a non-symmetric way by explicitly considering one process as source, the other as target variable. TE quantifies the part of information which is contained in the target variable and can be predicted by the source above and beyond the part of information which can be predicted by the past of the target variable [48]. In other words, TE determines the degree of uncertainty of one process which is exclusively due to another process. The discrete version of TE is defined as

$$TE_{Y\to X} = \sum_n p(x_{n+1}, x_n^{(m_x)}, y_n^{(m_y)}) \log \frac{p(x_{n+1}|x_n^{(m_x)}, y_n^{(m_y)})}{p(x_{n+1}|x_n^{(m_x)})} \tag{1}$$

with X as target, Y as source variable, x_n and y_n as the respective states at time point n, and m_x and m_y as the respective embedding dimensions (also called past or history length) of the variables. The embedding dimensions can be regarded as the assumed Markov orders of the processes. In this context, if a system can be approximated by a stationary Markov process of order k, then including the past of x_n beyond the k-th past time point x_{n-k+1} will not add any predictive information about the "future" state x_{n+1}. This Markov property, where each state in time depends only on the precedent states of the same process, is perturbed by the information flow from source to target process. Such deviation from Markov property is measured by TE. In respect of the definition above, we can see that the dynamic character of TE comes with the introduction of transition probabilities (terms in the logarithm) which allow to consider two systems at different time points. Henceforth, the term information transfer is used as underlying concept of TE, and therefore is always considered as directed information transfer.

TE have been used to study properties of time series data, especially in the field of medicine and neuroscience [49–51]. In this context, TE is largely considered as a measure for identifying direct information flow from one process to another, enabling causal inference [14, 43]. Before we continue, it is important to clarify the notion of causality regarding information transfer. If information transfer measured by TE should be a concept allowing causal inference, then causality must be considered in the sense of Granger causality [52]. This notion of causality is "based on the premise that cause precedes effect, and a cause contains information about the effect that is unique, and is in no other variable", as mentioned in [48]. Against this background, the intuitive view on "true" causality, which mostly implies some direct physical connection of the processes in question including the directionality of the impact, is not necessarily considered by Granger causality. Granger causality should be viewed as a measure for the directed statistical dependence of two processes. In this sense, TE and Granger causality share an inherently statistical nature, constructed to identify the predictive capacity of one process over another, and TE can be regarded "as a non-parametric test statistic for pure Granger causality" [48]. We have to note that for Gaussian processes, TE and Granger causality are identical quantities [48, 53], and this identity also holds for some cases of generalized normal distributions [54]. Regarding this work, we have to be cautious about causal inference, as for the goal of structural network analysis, we are interested in direct connections which is not necessarily equivalent to Granger causality. A measure to avoid false causal inferences is to apply a conditioned version of TE, which is introduced below.

Before applying information transfer analysis, we have to reflect that using TE on empirical data comes with limits and caveats regarding the sampling of the data, the dynamics of the system, the statistical aspects of TE estimation and the significance of information transfer.

As we assume continuous processes for biological examples, instead of discrete probability distributions, probability densities need to be estimated on discretized datasets via statistical methods. Therefore, the sampling of data has direct impact on the estimation of transition probabilities. Two important properties of the data should be considered: data length N and sampling frequency δt. In the literature, different approaches shed light on the influence of sampling upon TE estimation [55–57].

As Kaiser et al. emphasize, an important prerequisite for estimating TE is to assume stationary (or ergodic) dynamics, hence the probability distribution of the variables should remain relatively constant over the course of time [58]. Additionally, that work shows the importance of a convergent estimator of continuous entropy where the parameters of estimation strongly impact the quantification of information transfer. It is noteworthy

to mention that methods for estimating probability densities can be biased, introducing different types of statistical and systematic errors [58, 59]. Kernel density estimation is a non-parametric way to estimate probability density functions [60, 61], among which the use of a rectangular [18, 62] or Gaussian kernel [63] is a widespread approach. For this work, I applied an estimator using the k-nearest neighbor statistics developed by Kraskov, Stoegbauer, and Grassberger (KSG estimator) [64]. The advantage of the KSG estimator is that only a single parameter has to be set, that it exhibits high robustness with respect to variance of this parameter, and it is empirically reported to reduce the bias and variance of estimation [65]. However, we should note that there are no theoretical guarantees that the KSG estimator converges to the true underlying probability density [66, 67].
Additionally, it is crucial to validate the significance of TE estimations. Such validation is realized using surrogate-based data as adequate hypothesis tests. This non-parametric test method is based on creating new data through random permutations of the original time series [68]. Hence, the surrogate data is assumed to be uncorrelated noise which preserves the amplitude distribution of the original series, but destroys any correlation. In the context of information transfer, the original source variable is permutated and used as new surrogate-based source variable for TE estimation. Therefore, the hypothesis that a similar, but uncorrelated and noisy series exhibits comparable prediction power upon a target variable as the original source has to be rejected in order to validate the TE estimation with the original source-target constellation.

As we are interested in causal inference, let us consider different aspects of non-causal information transfer. For the notion of spurious information transfer, the connectivity of the processes X, Y and Z could be given by $X \leftarrow Z \rightarrow Y$, also referred to as common driver motif. Although X and Y are not connected through any path, the common source Z can induce spurious information transfer between the former processes, hence $TE_{X \leftarrow Y}$ is erroneously non-zero. If we examine the structure $X \rightarrow Z \rightarrow Y$, X and Y are connected via another process which induces indirect information transfer. Note that it is not erroneous to quantify a non-zero information transfer here, but it is incorrect to infer a direct causal relationship between X and Y due to non-zero $TE_{X \leftarrow Y}$.

This non-causal aspects of information transfer are the motivation for a conditioned form of TE. As Surovtsova et al. demonstrate, estimated TE values cannot readily be used for causal inference [69]. Depending on the structure and dynamics of the respective networks, non-causal TE can dominate the information transfer [70]. As reported by Smirnov, such non-causal information transfer can be due to the environment

of the observed processes which share common, but unobserved information sources, low temporal resolution and/or high noise level of the data [71].

In order to eliminate indirect information transfer and use TE as an inferrer of direct coupling, conditioned TE (CTE) has been applied by Surovtsova et al. CTE extends classical TE by conditioning other variables which are included in the network to analyze. Thus, otherwise unobserved processes, which are neglected in the bivariate constellation of unconditioned TE, are accounted for information transfer. Formally, the source variable X is conditioned on the target Y and environmental variable Z by

$$TE_{Y\to X|Z} = \sum_n p(x_{n+1}, x_n^{(m_x)}, y_n^{(m_y)}, z_n^{(m_z)}) \log \frac{p(x_{n+1}|x_n^{(m_x)}, y_n^{(m_y)}, z_n^{(m_z)})}{p(x_{n+1}|x_n^{(m_x)}, z_n^{(m_z)})} \tag{2}$$

which is analogues to Equation (1) with m_z denoting the environmental embedding dimension. A conditioned form of TE was first applied in [72], where only the target history length is variable and the source and environmental embedding dimensions are neglected, hence fixed to 1. Surovtsova et al.'s contribution to this topic is that all embedding dimensions are considered for the quantification of direct information transfer, and that analytical solutions of TE prove the possibility to eliminate spurious and indirect information transfer in the context of vector autoregression (VAR) models, if the embedding dimension are chosen correctly [69]. In that work, the authors show that a time-lagged variant of the source variable has to be included in the conditioning if one assumes that the source variable is affecting its own transition probabilities. In other words, if the source process Y contributes to the transition of $Y \to Y$, then the following definition of CTE should be adopted in order to eliminate non-causal information transfer:

$$TE_{Y\to X|ZY_{-1}} = \sum_n p(x_{n+1}, x_n^{(m_x)}, y_n^{(m_y)}, z_n^{(m_z)}, y_{n-1}^{(m_y)}) \log \frac{p(x_{n+1}|x_n^{(m_x)}, y_n^{(m_y)}, z_n^{(m_z)}, y_{n-1}^{(m_y)})}{p(x_{n+1}|x_n^{(m_x)}, z_n^{(m_z)}, y_{n-1}^{(m_y)})} \tag{3}$$

This definition of CTE is used in the framework of my thesis as I assume for biological systems that every process affects its own dynamics. Moreover, Surovtsova et al. show that non-causal information transfer can increase with increasing history lengths. Such observations can hint to the fact that unobserved processes act as additional information sources for the bivariate processes in question. Such behavior of information transfer will be referred to as information leak hereafter.

Note that in the context of causality detection and information flow [50], an additional time lag parameter enters the embedding procedure to optimally reconstruct the dynamics

of multidimensional processes [73, 74]. For this work, I will neglect the time lag as supposed by Schreiber and use the different history lengths to calculate information transfer [17].

In the literature, TE is not only considered as an inferrer of the existence of causal relations, but also as a quantifier of connectivity strength [63]. From this perspective, an interesting work by Chicharro et al. show that if one cannot assume the autonomy of the causing (sub-)system in comparison to the rest, interactions cannot be quantified in terms of cause and effect [75]. That results are generic which means that the validity of their conclusions only depends on the topology of the causal structure, not on the specific mechanism governing the dynamics. In this regard, highly interconnected systems, e.g. as the brain or some signal transduction networks, limit the use of approaches like Granger causality or TE for quantifying the connectivity strength. Although Chicharro et al. state the quantification of causal impact is not possible in the case of bidirectional causality (independent of the used measure), the qualification of causal connections can be determined under a strong restriction. For Granger causality, including TE, this restriction is the complete observability of the system, which also converges with the idea of a fully conditioned TE which takes account of all other processes included in the system. Hence, in the context of this work, the quantification of connectivity strength is not practicable, as the examined models are highly interconnected networks. In contrast, the inference of causality is in the focus of the thesis for which reason I will use the fully conditioned TE in order to suffice the restriction of complete observability.

In order to reveal the dynamics in networks, different methods have been applied using information transfer [43, 50, 56, 70]. Concerning biological signaling pathways, such analysis was performed upon Calcium signaling by Pahle et al. [18]. In [3, 37, 76, 77], different information-theoretic approaches to study signaling networks are reviewed. In this regard, MI was used to analyze the capacity of tumor necrosis factor (TNF) signaling to transduce information, focusing on bottleneck motifs [5]. Hormoz applies channel capacity and shows that the interdependent regulation of transcription factors (TFs) with overlapping transcription sites correlates the intrinsic noise in gene regulation, which maximizes the information flow in TF cross-regulation networks [78]. Another interesting framework for signal transduction is the rate distortion theory [79], a branch of information theory which enables to quantify optimal information encoding in the presence of noise [80]. This theory is used by Andrews et al. on chemotaxis to show that recurrent biological strategies like hysteresis or irreversibility induce optimal stimulus-response relations [81].

Regarding my thesis, I use information transfer to analyze the phosphorylation cascade of the MAPK signaling pathway. The goal is to assess the possibilities and limits of using TE to infer the structure of a highly interdependent and dynamic signaling network. In difference to other work, I assess the performance of conditioned TE upon biological signaling, especially the dependence of the TE estimation on different parameters, such as the embedding dimensions and temporal resolution, and how the choice of these parameters affects the inference of connectivity.
It is important to mention that deterministic simulation is inappropriate for TE estimations, as deterministically simulated time courses do not exhibit entropy in the information-theoretic sense. In other words, as the system is totally determined by the structure, kinetics, initial conditions and parameters, no information is generated or transferred inside such a system. Introducing random fluctuations into the system is a prerequisite for meaningful TE estimations. This can be achieved by using stochastic simulation methods, which are introduced in Section 1.3.

1.2. MAPK Signaling Pathway

The Mitogen Activated Protein Kinases (MAPKs) are an important phosphorylation cascade which are recurrent building blocks in various signaling pathways that regulate cell functions such as metabolism, differentiation, division, growth and apoptosis, among others [19, 20, 82]. The structure of the cascade as well as the protein domains are evolutionary conserved in all eukaryotic cells and consist of three layers of protein serine threonine kinases which respond to extracellular stimuli and trigger different cellular responses by sequentially activating each other in the following order: MAP kinase kinase kinase (M3K), MAP kinase kinase (M2K), and MAP kinase (MK) [20]. The terminal layer of MKs is activated via upstream M2Ks at two phosphorylation sites, conserved threonine and tyrosine residues. In turn, M2Ks are phosphorylated at serine and threonine residues by M3Ks [20]. These two levels necessitate double phosphorylation for activation. The first level of M3Ks can be activated via different mechanisms (interaction with GTP-binding proteins, oligomerization, subcellular relocalization) which can involve phosphorylation of a tyrosine residue, e.g. in the case of Raf. Phosphotases can inactivate the corresponding kinases at each cascade level. An extensive overview about the activation and function of MAPKs can be found in [20].

The series of covalent modification cycles of kinase cascades can function as an amplifier of the signal, converting a small number of signaling molecules to a high number of target molecules. In the specific case of MAPK, an important function of the cascade is increasing the sensitivity regarding the stimulus, where each additional level of the phosphorylation cascade multiplies the signal response [83]. The (de-)phosphorylation reactions possess regulatory function by integrating negative and positive feedback loops in the cascade [84]. It is well known that system dynamics can be determined via feedback loops. Specifically positive feedback can induce switch-like or bistable behavior [85], whereas negative feedback can increase the stability of the system and induce oscillatory behavior [86]. For the MAPK cascade, it is shown that the interplay of negative and positive feedback loops shapes the system dynamics of signal transduction [84]. That study also illustrates that feedback loops not only induce oscillatory behavior, but also modulate the amplitude and frequency of the oscillations, and therefore determine the duration and robustness of the response. In this regard, the design of feedback loops can be crucial in order to preserve certain properties of the oscillation upon variance of the signal or cellular components [87]. Past research suggests that control systems encode biologically relevant information in certain properties of their output in order to determine cell fate specifically [76]. Particularly oscillatory behavior can encode such relevant information in the amplitude and frequency [88–90]. Therefore, feedback loops can be crucial for inducing robust encoding behavior regarding specific properties of the input, while being sensitive to other forms of signal variation at the same time. This stimuli dependent robustness and sensitivity of the response is particularly observable for oscillatory behavior [87]. On the other hand, the information content of oscillations can be richer, increasing the potential of differential decoding of the signal for downstream processes [90].

The ERK pathway is a prominent signal transduction pathway which involves the MAPK subsystem (or the Raf-MEK-ERK pathway) to regulate cell proliferation, survival and differentiation [91]. Its phosphorylation cascade including different feedback loops has been identified as an important element for coordinating the dynamic behavior of signal transduction and determines cellular outcome [92]. Nonetheless, the dynamic control and plasticity of MAPKs including signaling pathways is not fully understood as the dynamic complexities need elaborate (computational, mathematical) tools for analyzing the systems-level properties of the regulatory modules [89].

In the context of information transfer analysis, a simplified version of the MAPK cascade is used, which traces back to the modelling approach of Kholodenko [86]. In this version, several possible feedback and feedforward loops are not integrated, which are present

in more elaborated and current models (for these models, see [87]). As the MAPK model should serve as a test case for TE analysis, a less interdependent structure is favored. Nevertheless, the underlying kinetics and topology of the used MAPK model are biologically adequate and identical to the more complex models.

1.3. Stochastic Simulation

Deterministic methods are the most widespread approach for simulating models. This methods are based on numerical integration of ordinary differential equations (ODEs) which define the reactions of a (bio-)chemical system. Deterministic simulation has a couple of advantages such as fast computation, well defined outcome production, a variety of analytical methods and abundant simulation algorithms [93, 94]. On the other side, deterministic simulations can be insufficient to display certain system dynamics which unfold via stochastic effects. Stochastic stimulation can be used to introduce such random fluctuations into the system. Dynamical systems can exhibit multi-stable outcomes, which means that given certain initial conditions, the system can evolve towards multiple numerical states. Such multi-stability cannot be reflected by the deterministic approach, which leads to a single-stable outcome where all other possible results are canceled out. In contrast, alternate outcomes can be adequately represented by means of stochastic simulation. The Lotka-Volterra predator-prey model is an example for such a multi-stable system. For systems with low particle numbers, the deterministic approach can produce unrealistic outputs failing to display dynamics due to random fluctuations. Regarding (bio-)chemical reactions, the order and exact time points when certain reactions will occur cannot be determined. Reactions are rather governed by likelihoods which allows to determine the probability of occurrence of one reaction compared to another in a certain time interval. Particularly for systems with low abundant (bio-)chemical species, random fluctuations can have a huge impact on the results of the simulation. Such qualitative difference in the outcome due to randomness in the system can be observed in the Lotka-Volterra model where the extinction or survival of a species are possible consequences [95]. Stochastic algorithms, such as Gillespie's Direct Methods, account for random fluctuations and hence are adequate approaches for simulating certain models [96].

Mainly, stochastic simulation can be useful for cases where random fluctuations possess biological significance. The lysis or lysogeny pathway of λ-infected *Escherichia coli* cells

is an example where the fluctuation of gene expression is essential for cell fate decision [97]. In the presence of molecular noise, cells exhibit mechanisms to induce robustness against random perturbations [98, 99], or noise can even have a positive impact on biological systems [100].
In the context of my thesis, a stochastic method is used to simulate the MAPK model in order to perform a TE analysis for two reasons. First, as the MAPK model is a signaling pathway, the species are not present in high abundance. Therefore random fluctuations should be taken into account. Second, information transfer analysis requires sources of information upon which TE estimation can be performed. As mentioned in Subsection 1.1.2, this is not given for deterministic simulations. In contrast, random fluctuations can introduce entropy into the system by what a meaningful information transfer analysis can be realized.

Most analytical methods, which are available for the deterministic approach, such as Metabolic Control Analysis (MCA) and sensitivity analysis, are not applicable for stochastic simulation [101, 102]. Additionally, the computational effort can be huge as the simulation time roughly scales with number of particles, hence, even the simulation of simple models can take up a lot of time. Therefore, whether stochastic or deterministic simulations should be applied depends on the dynamical properties of the system and the questions to be investigated.

For reasons of completeness, I also mention the hybrid methods, which combine the advantages of both approaches [103]. Note that such methods are not used in the framework of my thesis. Hybrid methods subdivide the system into different parts and adequate methods operate on these parts simultaneously. As most fast reactions involve high abundant species (e.g. metabolism), and most slow reactions concern low-particle species (e.g. gene expression, signal transduction), the partitioning in fast and slow subsystems allows to apply numerical integration of ODEs and stochastic simulations respectively. For an extensive review of stochastic simulation and different simulation methods, refer to [104].

1.4. Structure of the Thesis

In the following parts of the thesis, a TE analysis of connectivity is employed upon the MAPK model. But before I perform such an analysis, I created different simple

models as generic test cases for assessing the reliability and robustness of TE estimations. The simple models come in two variants: linear and biological models. The simple linear models exhibit mass actions as underlying kinetics and are referred to as SL henceforth. The SL models are the attempts of translating from a linear VAR process with three variables to a corresponding system of biochemical reactions which exhibit biologically plausible connections. The term linear here refers to the underlying kinetics of the SL models (see Subsec. 2.2.1). As Surovtsova et al. studied CTE via linear VAR processes, the creation of the SL models is motivated to follow that research [69]. In contrast, the biological models exhibit biologically plausible kinetics and connectivity, where the underlying mechanisms are simple Michaelis-Menten kinetics and the topology exhibits circular structure. These structures are supposed to mimic the feedback motif in biological signal transduction, which renders the network highly interdependent. Two different variants of simple biological models are present: generic biological models and a biological model which is a minimized version of the MAPK model, which are hereafter referred to as SB and SM respectively. In the case of SM, the model contains 3 motifs which are vital for the MAPK cascade: A (de-)phosphorylation cycle of a single species, a phosphorylation cascade (where a phosphorylated species enhances the phosphorylation of another species) and a feedback regulation to an upstream reaction.
TE analysis is performed by advancing from the simplest model to the more complex variants, hence in the following order: SL, SB, SM and MAPK models. The estimation of information transfer is evaluated upon changing the parameters concerning the data and estimation process, such as the data length, sampling frequency and embedding dimensions. The procedure of scanning different data lengths and sampling frequencies is referred to as N-scan and δt-scan respectively. In Chapter 4, a detailed discussion about the results of Chapter 3 are given. A possible best practice procedure for information transfer analysis and an outlook for future research are presented in Sections 4.7 and 4.8 respectively.

2. Materials and Methods

2.1. Stochastic Simulation Algorithm

The stochastic simulation was performed using the Direct Method developed by Gillespie [96]. Gillespie approximated biochemical systems as colluding particles leading to possible reactions, calculating average probabilities for these events. Within this framework, the so-called Chemical Master Equation can be formulated which defines the temporal state probability distribution of the system, given the initial state of the species x_0 at the point in time t_o [96]:

$$\frac{\delta P(x,t|x_0,t_0)}{\delta t} = \sum_{\mu=1}^{M} a_\mu(x-v_\mu)P(x-v_\mu,t|x_0,t_0) - a_\mu(x)P(x,t|x_0,t_0)$$

R_μ represents a single reaction in the system, x denotes the vector of species particle numbers, v_μ the stoichiometric vector of reaction R_μ, and a_μ the propensity of R_μ. Propensities can be understood as tendency to yield a reaction event and are derived from the deterministic reaction rates [104]. As this Chemical Master Equation can be solved analytically or numerically only for limited cases, a method which simulates the underlying stochastic process has to be applied. Such stochastic simulation algorithms can be divided into exact, approximate and hybrid methods, while only an exact method was used and will be introduced as concept.

Exact methods explicitly simulate each reaction in the system, therefore correctly account for random fluctuations and the discrete nature of the system. Hence, these methods are valid even for systems with very small particle numbers. In general, exact methods are formally described by the so-called Reaction Probability Density Function:

$$P(\tau, \mu|x, t) = a_\mu(x) \exp\left(- \sum_{\mu=1}^{M} a_\mu(x)\tau \right)$$

Given the state x_μ and at time t_μ, it determines the probability density that a reaction R_μ will occur after the time step τ. It represents a homogeneous Poisson process and assumes that the propensities are conserved upon time evolution. The system can be simulated by drawing random numbers according to that density function, triggering the concerned reactions, updating the state of the system and progressing in time, by iteratively realizing the reactions.
For the Direct Method, two uniformly distributed random numbers have to be generated: the first one (τ) defines the stochastic time step when the next reaction will occur, and the second one determines the type of reaction R_τ to occur. For every reaction, the according propensities are calculated and the sum a_0 over all propensities a_μ is used to calculate the stochastic time step τ:

$$\tau = -\frac{1}{a_0} \ln \tau_1$$

with τ_1 denoting an uniformly distributed random number between 0 and 1 excluding 0. The Monte Carlo step finally determines which reaction event occurs, by generating the second random number and, with that, "drawing" a reaction R_τ with a probability proportional to the relative propensity a_τ/a_0, which lies in the range of the random number. Thus, the probability to trigger reaction R_τ is proportional to the propensity a_τ. After realizing the reaction, the number of particles is updated according to the stoichiometry and the time is incremented by τ. The process is repeated until the defined simulation time is reached.

2.2. Models

2.2.1. Simple Linear Models

The simple linear (SL) models consist of 3 species which are interconnected in a triangular structure. The underlying kinetics governing the dynamics of the system are mass actions

and constant influx reactions. Both models were set to a total volume of 10^{-18} ml. The reactions and kinetics of the models SL1 and SL1.2 are given in Table 1. The simulations for model SL1 were performed with the initial concentrations $[A]_0, [B]_0, [C]_0 = 2$ [a.u.], and with $[A]_0, [B]_0, [C]_0 = 0.1$ [a.u.] for model SL1.2.

Table 1: Reactions and kinetics of the simple linear models SL1 and SL1.2. Parameters of model SL1: $k_1, k_2, k_3, k_4, k_5, k_6 = 0.1$, $k_7, k_8, k_9 = 0.2$. Parameters of model SL1.2 differing from model SL1: $k_4, k_5, k_6 = 0.5$, $k_7, k_8, k_9 = 1$.

reactions			kinetics
	$\rightarrow$	A	$v_1 = k_1$
	$\rightarrow$	B	$v_2 = k_2$
	$\rightarrow$	C	$v_3 = k_3$
	$\xrightarrow{C}$	A	$v_4 = k_4 \cdot [C]$
	$\xrightarrow{A}$	B	$v_5 = k_5 \cdot [A]$
	$\xrightarrow{B}$	C	$v_6 = k_6 \cdot [B]$
A	$\rightarrow$		$v_7 = k_7 \cdot [A]$
B	$\rightarrow$		$v_8 = k_8 \cdot [B]$
C	$\rightarrow$		$v_9 = k_9 \cdot [C]$

2.2.2. Simple Biological Models

For the simple biological models, Michaelis-Menten kinetics were used, and 3 different underlying structures are given. The models SB1 and SB1.2 consist of 3 species where circular interconversion occurs (Table 2).
The models SB2 and SB2.2 consist of 6 species where every moiety shows an activated and inactivated species. The activation reactions are modulated by another moiety (modifier), and interconversions between the species occurs (Table 3).
Model SM consists of 4 species where every moiety shows a "phosphorylated" and "unphosphorylated" species. The phosphorylation reactions are modulated by the other moiety (Table 4).
Model SM mimics 3 motifs which are vital for the MAPK cascade: A (de-)phosphorylation cycle, a phosphorylation cascade and a negative feedback regulation. All simple biological models were set to a total volume of 10^{-18} ml. The simulations for the models SB1 and SB1.2 were performed with the initial concentrations $[A]_0, [B]_0, [C]_0 = 2$ [a.u.]. For the models SB2 and SB2.2, the initial concentration are $[A]_0, [B]_0, [C]_0 = 0.1$ [a.u.], $[A_i]_0, [B_i]_0, [C_i]_0 = 0.5$ [a.u.]. For model SM, the initial concentrations are set to $[A_i]_0, [B_i]_0, [C_i]_0 = 1$ [a.u.].

Table 2: Reactions and kinetics of the simple biological models SB1 and SB1.2. Parameters of model SB1: $k_1, k_2, k_3 = 0$, $k_4, k_5, k_6 = 0.025$, $K_4, K_5, K_6 = 15$, $k_7, k_8, k_9 = 0$. Parameters of model SB1.2 differing from model SB1: $k_1, k_2, k_3 = 1$, $k_7, k_8, k_9 = 0.5$.

reactions			kinetics
	$\rightarrow$	A	$v_1 = k_1$
	$\rightarrow$	B	$v_2 = k_2$
	$\rightarrow$	C	$v_3 = k_3$
A	$\rightarrow$	B	$v_4 = \frac{k_4 \cdot [\mathrm{A}]}{K_4 + [\mathrm{A}]}$
B	$\rightarrow$	C	$v_5 = \frac{k_5 \cdot [\mathrm{B}]}{K_5 + [\mathrm{B}]}$
C	$\rightarrow$	A	$v_6 = \frac{k_5 \cdot [\mathrm{C}]}{K_5 + [\mathrm{C}]}$
A	$\rightarrow$		$v_7 = k_7 \cdot [\mathrm{A}]$
B	$\rightarrow$		$v_8 = k_8 \cdot [\mathrm{B}]$
C	$\rightarrow$		$v_9 = k_9 \cdot [\mathrm{C}]$

Table 3: Reactions and kinetics of the simple biological models SB2 and SB2.2. Parameters of model SB2: $k_1, k_2, k_3 = 0.01$, $k_4, k_5, k_6 = 1$, $K_4, K_5, K_6 = 5$, $k_7, k_8, k_9 = 0.02$, $k_{10}, k_{11}, k_{12} = 0$. Parameters of model SB2.2 differing from model SB2: $k_4, k_5, k_6 = 5$, $K_4, K_5, K_6 = 7$, $k_7, k_8, k_9 = 0.15$, $k_{10}, k_{11}, k_{12} = 0.01$.

reactions			kinetics
	$\rightarrow$	$\mathrm{A_i}$	$v_1 = k_1$
	$\rightarrow$	$\mathrm{B_i}$	$v_2 = k_2$
	$\rightarrow$	$\mathrm{C_i}$	$v_3 = k_3$
$\mathrm{A_i}$	$\overset{C}{\rightarrow}$	A	$v_4 = \frac{k_4 \cdot [\mathrm{A_i}] \cdot [\mathrm{C}]}{K_4 + [\mathrm{A_i}]}$
$\mathrm{B_i}$	$\overset{A}{\rightarrow}$	B	$v_5 = \frac{k_5 \cdot [\mathrm{B_i}] \cdot [\mathrm{A}]}{K_5 + [\mathrm{B_i}]}$
$\mathrm{C_i}$	$\overset{B}{\rightarrow}$	C	$v_6 = \frac{k_6 \cdot [\mathrm{C_i}] \cdot [\mathrm{B}]}{K_6 + [\mathrm{C_i}]}$
A	$\rightarrow$		$v_7 = k_7 \cdot [\mathrm{A}]$
B	$\rightarrow$		$v_8 = k_8 \cdot [\mathrm{B}]$
C	$\rightarrow$		$v_9 = k_9 \cdot [\mathrm{C}]$
$\mathrm{A_i}$	$\rightarrow$		$v_{10} = k_{10} \cdot [\mathrm{A_i}]$
$\mathrm{B_i}$	$\rightarrow$		$v_{11} = k_{11} \cdot [\mathrm{B_i}]$
$\mathrm{C_i}$	$\rightarrow$		$v_{12} = k_{12} \cdot [\mathrm{C_i}]$

Table 4: Reactions and kinetics of the simple biological model SM. Parameters: $k_1, k_2, k_3, k_4 = 0.01$, $K_1, K_2, K_3, K_4 = 15$, $K_i = 10$, $n = 1$.

reactions			kinetics
A	$\xrightarrow{\mathrm{B_p}}$	$\mathrm{A_p}$	$v_1 = \frac{k_1 \cdot [\mathrm{A}]}{(K_1 + [\mathrm{A}]) \cdot \left(1 + \left(\frac{[\mathrm{B_p}]}{K_i}\right)^n\right)}$
B	$\xrightarrow{\mathrm{A_p}}$	$\mathrm{B_p}$	$v_2 = \frac{k_2 \cdot [\mathrm{B}] \cdot [\mathrm{A_p}]}{K_2 + [\mathrm{B}]}$
$\mathrm{A_p}$	$\rightarrow$	A	$v_3 = \frac{k_3 \cdot [\mathrm{A_p}]}{K_3 + [\mathrm{A_p}]}$
$\mathrm{B_p}$	$\rightarrow$	B	$v_4 = \frac{k_4 \cdot [\mathrm{B_p}]}{K_4 + [\mathrm{B_p}]}$

2.2.3. MAPK Models

The details about the structure of the MAPK cascade can be found in Section 1.2. It is important to emphasize that 3 recurrent biological motifs can be found in the MAPK model: (de-)phosphorylation cycles, a phosphorylation cascade and a feedback loop. The reactions and kinetics of the models M1 and M2 are given in Table 5, where the negative feedback regulation from $\mathrm{MK_{pp}}$ to the phosphorylation reaction of M3K (reaction v_1) shows an additional inhibition term. Otherwise, the model is based on (modified) Michaelis-Menten kinetics. Varying parameter K_i results in qualitative changes in the dynamics of the MAPK model, switching from steady-state dynamics to oscillatory behavior [86]. M1 and M2 were set to a total volume of 10^{-14} ml. The simulations for both models were performed with following initial concentrations: $[\mathrm{M3K}]_0 = 90$, $[\mathrm{M3K_p}]_0 = 10$, $[\mathrm{M2K}]_0 = 280$, $[\mathrm{M2K_p}]_0 = 10$, $[\mathrm{M2K_{pp}}]_0 = 10$, $[\mathrm{MK}]_0 = 280$, $[\mathrm{MK_p}]_0 = 10$, $[\mathrm{MK_{pp}}]_0 = 10$ (given in [a.u.]).

2.3. Transfer Entropy Estimation

The main script for estimating TE using the method described by [64] was written by Irina Surovtsova (Bioquant, Heidelberg). I made changes to Surovtsova's original script which prepares the data structure for TE estimation. These changes enabled fully conditioned TE estimation, as the original script only allowed for the conditioning of the source process, not for environmental processes. My script (see `TE_KraskovData-mod.R` in `https://github.com/trotmano/Master-thesis`) creates a data structure according to the type of TE estimation which should be performed: no conditioning (Eq. (1)), conditioning of the source process and conditioning of the source and environmental

Table 5: Reactions and kinetics of the MAPK models M1 and M2. Parameters for M1: $k_1 = 2.5$, $k_2 = 0.25$, $k_3, k_4, k_7, k_8 = 0.025$, $k_5, k_6 = 0.75$, $k_9, k_{10} = 0.5$, $K_1 = 10$, $K_2 = 8$, $K_8 = 30$, $K_3, K_4, K_5, K_6, K_7, K_9, K_{10} = 15$, $K_i = 9$, $n = 1$. Parameters for M2 differing from M1: $K_i = 23$.

reactions			kinetics
$\mathrm{M3K}$	$\xrightarrow{\mathrm{MK_{pp}}}$	$\mathrm{M3K_p}$	$v_1 = \frac{k_1 \cdot [\mathrm{M3K}]}{(K_1 + [\mathrm{M3K}]) \cdot \left(1 + \left(\frac{[\mathrm{MK_{pp}}]}{K_i}\right)^n\right)}$
$\mathrm{M3K_p}$	$\rightarrow$	$\mathrm{M3K}$	$v_2 = \frac{k_2 \cdot [\mathrm{M3K_p}]}{K_2 + [\mathrm{M3K_p}]}$
$\mathrm{M2K}$	$\xrightarrow{\mathrm{M3K_p}}$	$\mathrm{M2K_p}$	$v_3 = \frac{k_3 \cdot [\mathrm{M2K}] \cdot [\mathrm{M3K_p}]}{K_3 + [\mathrm{M2K}]}$
$\mathrm{M2K_p}$	$\xrightarrow{\mathrm{M3K_p}}$	$\mathrm{M2K_{pp}}$	$v_4 = \frac{k_4 \cdot [\mathrm{M2K_p}] \cdot [\mathrm{M3K_p}]}{K_4 + [\mathrm{M2K_p}]}$
$\mathrm{M2K_{pp}}$	$\rightarrow$	$\mathrm{M2K_p}$	$v_5 = \frac{k_5 \cdot [\mathrm{M2K_{pp}}]}{K_5 + [\mathrm{M2K_{pp}}]}$
$\mathrm{M2K_p}$	$\rightarrow$	$\mathrm{M2K}$	$v_6 = \frac{k_6 \cdot [\mathrm{M2K_p}]}{K_6 + [\mathrm{M2K_p}]}$
MK	$\xrightarrow{\mathrm{M2K_{pp}}}$	$\mathrm{MK_p}$	$v_7 = \frac{k_7 \cdot [\mathrm{MK}] \cdot [\mathrm{M2K_{pp}}]}{K_7 + [\mathrm{MK}]}$
$\mathrm{MK_p}$	$\xrightarrow{\mathrm{M2K_{pp}}}$	$\mathrm{MK_{pp}}$	$v_8 = \frac{k_8 \cdot [\mathrm{MK_p}] \cdot [\mathrm{M2K_{pp}}]}{K_8 + [\mathrm{MK_p}]}$
$\mathrm{MK_{pp}}$	$\rightarrow$	$\mathrm{MK_p}$	$v_9 = \frac{k_9 \cdot [\mathrm{MK_{pp}}]}{K_9 + [\mathrm{MK_{pp}}]}$
$\mathrm{MK_p}$	$\rightarrow$	MK	$v_{10} = \frac{k_{10} \cdot [\mathrm{MK_p}]}{K_{10} + [\mathrm{MK_p}]}$

processes (Eq. (3)). Also, the choice of the history lengths m_x, m_y and m_z are taken into account. Subsequently, this data structure is handed over to the original script written by Surovtsova which performs the TE estimation.

Before quantifying the information transfer, the data has to prepared for TE estimation. A small amount of uniformly distributed noise (minimum and maximum 10^{-9}) was added to the data to prevent numerically identical values. Subsequently, the data was normalized (zero mean and unit variance) prior to TE estimation.
The TE estimation procedure is adopted according to Schreiber [17]. The general concept of TE is introduced in Subsection 1.1.2 and given in Equation (3). Here, I concisely describe the KSG estimation algorithm developed by Kraskov et al. [64]. The KSG algorithm is based on k-nearest neighbor statistics. In this methods, the algorithm searches for the k-th nearest neighbor around a given data point in the highest dimensionality of the probability distributions to approximate. The density of the so constructed box around the given data is assumed to be constant. Based on this, the local density around the given data point is estimated, and the algorithm proceeds to the next data point. The benefit of this algorithm is that only one parameter have to

be set, and that it is reported to be stable and displays low estimator bias [65]. But as mentioned above, no theoretical guarantees exist that the estimator converges to the true underlying probability density with infinite number of samples [66, 67]

For the validation of the TE estimations, a surrogate based test method was performed using permuted data [17]. The permutation of the source process yields a random sequence with identical probability distribution as the original data. The specific temporal relations between source and target process should be disrupted, as the temporally specific contributions of the source variable to the transition probabilities of the target are perturbed (see Subsec. 1.1.2). Hence, the TE estimation performed upon surrogate data exhibits the bias in information transfer which can be quantified only due to the probability distributions displayed by the original data, not the specific temporal correlations. In order to reject the the null hypothesis that a random sequence of the given source data can be used in the same degree to predict the target process as the original data, significant difference between surrogate and original estimates have to detected (see Sec. 2.4 for further details).

As the information transfer depends strongly on the choice of the history lengths, the TE estimations were performed for a range of different m_x, m_y and m_z values. In order to assess the TE estimations upon varying number of data points (data length) N and the interval time between two data points (sampling frequency) δt, these two sampling parameters were scanned in specific ranges adapted to the used models. The scanning procedures are referred to as N-scan and δt-scan respectively.

2.4. Data Analysis

Statistical testing was performed using the one-sample Wilcoxon signed rank test with a significance level of $p \leqslant 0.01$ [105]. More precisely, the TE value of a single original estimate was compared to 1000 corresponding surrogate estimates. Hence, with a probability of $p \leqslant 0.01$, one cannot reject the null hypothesis that the original TE value is the median of the given surrogate probability distribution. The Wilcoxon signed rank test was chosen as normality of the surrogate estimate distributions was not given in every case. As multiple comparison were performed (20 original estimates upon a specific TE estimation parametrization), the familiy-wise error rate has to be controlled [106]. For this purpose, Holm-Bonferroni correction was applied [107]. Note that such

statistical testing was only applied for certain models and δt values in the δt-scan procedure. Only for δt values which did not exhibit information leak behavior such validation was performed. Otherwise, for every single TE estimation parametrization, 10 original estimates were performed, and for every of these original estimates 10 surrogate estimations were calculated.

2.5. Used Software and Hardware

Models were constructed using COPASI (version 4.24, build 197 for 86_64 GNU/Linux) [108]. The different scanning procedures, TE estimations and simulations were executed via R script (version 3.5.2) [109] coupled to COPASI binaries using the CoRC package (version 0.4.0) [110]. The Direct Method implemented in COPASI was used for stochastic simulation. The calculations were performed on linux (CentOS 6.4 x86_64Kernel 2.6.32) compute cluster with Sun Fire AMD Opteron and IBM Quad Intel Xeon CPUs. The results were transferred back to R and processed, analyzed and plotted using the packages tidyr, plyr, dplyr and ggplot2 [111–114]. R scripts, models and an electronic version of the thesis can be found in `https://github.com/trotmano/Master-thesis`.

3. Results

One of the most important parameters for TE estimation is the sampling frequency of the data. Some dynamics of the system can be reflected or concealed in the structure of the data, depending on the temporal resolution. As information transfer analysis depends on the successful estimation of transition probabilities, too high sampling frequencies can mask the information transfer between certain processes, as decreasing δt values can result in a general decrease of information transfer up to a level of insignificant TE. In contrast, too low sampling frequencies can result in the loss of information about the dynamics. These tendencies were reported by Schoch [57]. In order to find suitable sampling frequencies for a TE analysis which allows inference to causal structures, δt-scans with variable embedding dimensions were performed upon different models. Additionally, for some cases the dependency between information transfer and the data length is assessed in order to ensure that the number of data points used for TE estimation is sufficient for a inference of the underlying connectivity of the system.

3.1. Simple Model

3.1.1. Linear Kinetics

The simple linear models SL1 and SL1.2 are the attempts of translating from a linear VAR process with three variables (and discrete time steps) to a corresponding system of biochemical reactions which exhibit biologically plausible connections (see Subsec. 2.2.1 and Fig. 1). As linear VAR processes have been successfully analyzed with information transfer to infer causality [69], I used this corresponding system to test the capability of TE estimations to reveal the connectivity of a network based on differential equations. For model SL1, we can see that the TE estimation depends strongly on the sampling frequency (Fig. 2). For some δt values, the information transfer increases with increasing history length, for other δt values the opposite holds. A convergence of TE dependent on

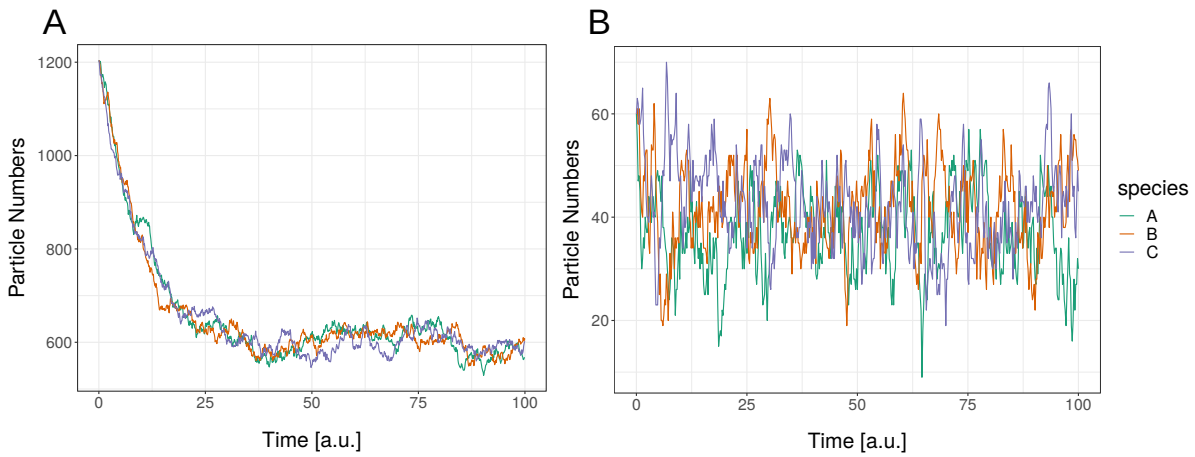

Figure 1: Time courses of the simple linear models. Time series for model SL1 **(A)** and SL1.2 **(B)** were generated using Direct Method as stochastic simulation algorithm. Both models reach a steady-state. The shown instances were simulated with the sampling frequencies 0.1 and 0.2 [a.u.] for model SL1 and SL1.2 respectively.

the sampling frequency is not observable, therefore determining the range of adequate δt values in the domain of convergence becomes impractical. Furthermore, choosing δt on the ground of maximal TE values is also not a suitable method, as the information transfer in this range increases with the increasing history lengths. Such behavior, also referred to as information leak, can indicate that non-causal TE is present regarding two coupled processes. In other words, TE also quantifies a portion of indirect and/or spurious information transfer between two processes which can be a result of incorrect parametrization and/or conditioning of the TE estimation. But it cannot be excluded that certain sampling frequencies induce statistical errors in the TE estimation process, yielding numerical instability. Besides, we can observe that the maximal TE values do not reflect the underlying connectivity in the model at all, exhibiting similar information transfer levels for all couplings independent of the choice of target or source variable. Hence, other criteria than convergence or maximal information transfer have to be taken into account to determine suitable sampling frequencies for a TE analysis of connectivity. The identification of such criteria will be evolved based on the results of the simple linear model.

In the following subsection, only the fully conditioned CTE is considered for further analysis, and hence, the environmental history length is simply referred to as history length if not indicated otherwise. For model SL1, we can identify sampling frequencies ($\delta t = [2, 8]$ [a.u]) for which the information transfer reflects the underlying structure of the system (Fig. 2). If the sampling frequency is too high, the information transfer increases with increasing history length and the surrogate estimates show negative mean values. For lower sampling frequencies, the surrogate estimates are distributed around

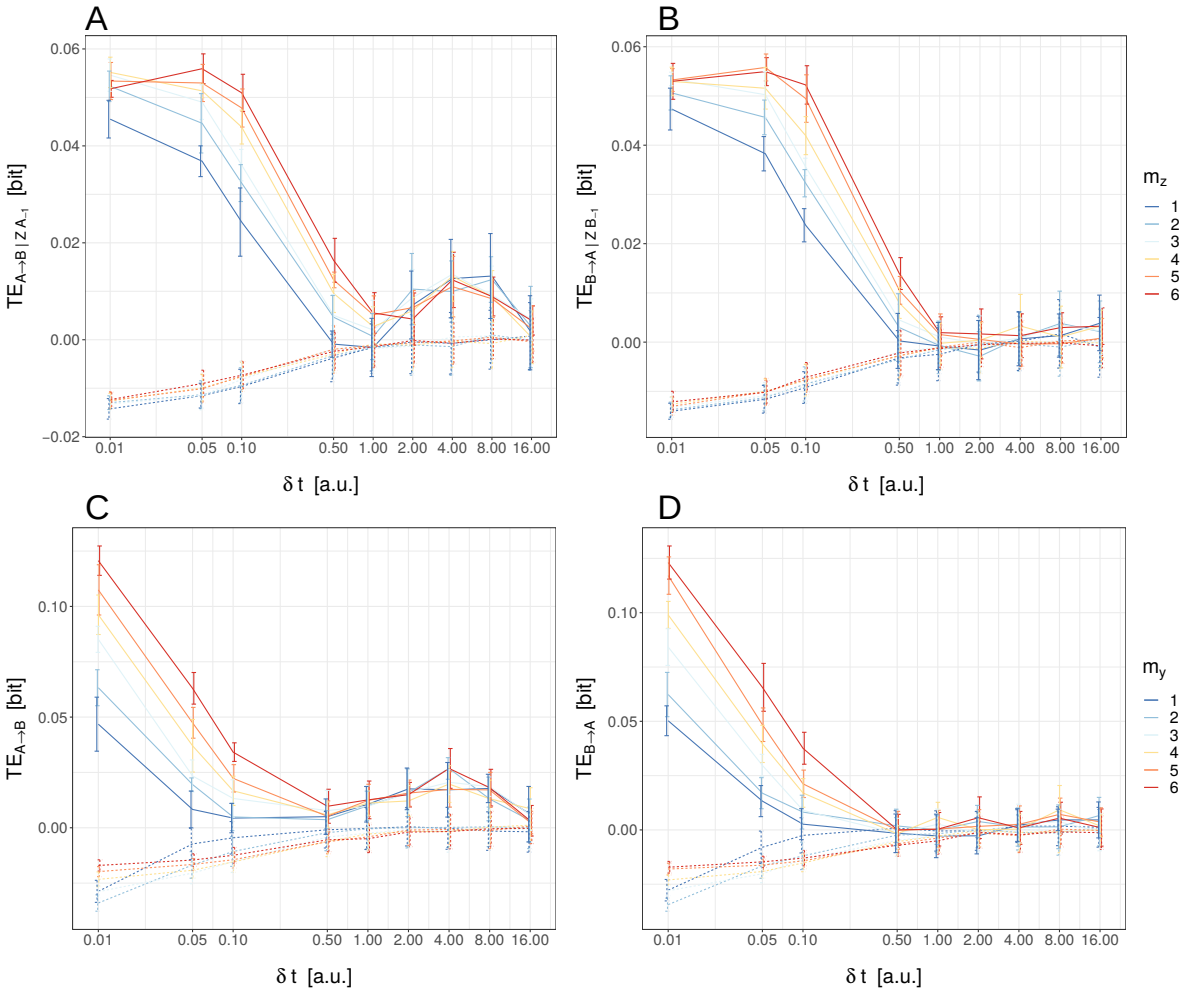

Figure 2: δt-scan for model SL1. TE estimations in dependency of the sampling frequency δt with varying environmental and source history lengths m_z and m_y respectively are shown. Information transfer for different source-target constellations between the species of model SL1 are computed: **A,C:** $A\to B$, **B,D:** $B\to A$. Among these, only $A\to B$ is a truly connected coupling. The upper part displays conditioned, the lower part unconditioned TE. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original and 100 surrogate estimates respectively). Computations are realized with data length $N = 5000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.

0 bits, and the information transfer is preserved upon changing the history length, or slightly decreases with increasing past length. In this range, the difference between original and surrogate estimates is lower compared to high sampling frequencies, but the information transfer reflects the underlying connectivity in the model. At a certain point, too low sampling frequencies lead to insignificant TE estimations ($\delta t = 16$ [a.u]). Thus, considering the δt-scan of model SL1, we can observe two phenomena which coincide with the range of sampling frequencies which allow us to use the information transfer as inferrer of causal structures: first, with increasing history length, CTE does not increase for the original estimates, and second, the surrogate estimates are distributed

around 0 bits, independent of the choice of history length. It can be argued that the second observation cannot be taken as a criterion for determining adequate sampling frequencies as the KSG estimator is not a kernel density estimator, hence estimated probability distributions do not unconditionally sum to unity. Thereby, the TE estimation using the KSG method can numerically exhibit negative values. But nonetheless, in the case of SL1, we can observe that the second criterion strongly coincides with the range of sampling frequencies which can be identified as suitable for TE analysis of connectivity. In contrast, the first observation is suitable as criterion as it is an indicator of information leak which should be reduced as much as possible for the goal of inferring causality. This criterion should hold for conditioned, but no necessarily for unconditioned TE.

Generally, unconditioned TE estimations exhibit higher values compared to the conditioned form. Directionally non-coupled species do not show any information transfer with unconditioned TE. As the unconditioned form measures causal as well as non-causal information transfer, regarding the closed circle structure and strong interdependence of the model, we would expect that every species appears as significant information source for any other target in the system. But in this case, for uncoupled source-target constellations, the information transfer is similar using unconditioned and conditioned TE. Additionally, the range of sampling frequencies which allow a TE analysis of connectivity depend on the structure and dynamics of the system. Another alternative model (SL1.2), analogous to the model structure of SL1 (see Subsec. 2.2.2), but with slightly modified connectivity and kinetics, shows that this range of sample frequencies is strongly model specific (see App. A Fig. A.1). In that case, the sampling frequencies which do not exhibit information leak behavior lie between 0.5 – 4 [a.u.], which is a significant difference to the range observed in model SL1. Therefore, even small changes in the structure and dynamics of a model impact the outcome of TE estimations.

3.1.2. Biological Kinetics

For the purpose of examining the performance of TE analysis upon more biological kinetics and connections, minimal models were generated where the underlying mechanisms are simple Michaelis-Menten kinetics and the topology exhibits circular structure (see Subsec. 2.2.2 and Fig. 3). This structure is also supposed to mimic the feedback motif in biological signal transduction, rendering the network highly interdependent. Model SB1 is a simple system which forms a closed cycle of 3 mass conversions between 3 species. Of course, such closed conversions are biochemically not plausible, but in this

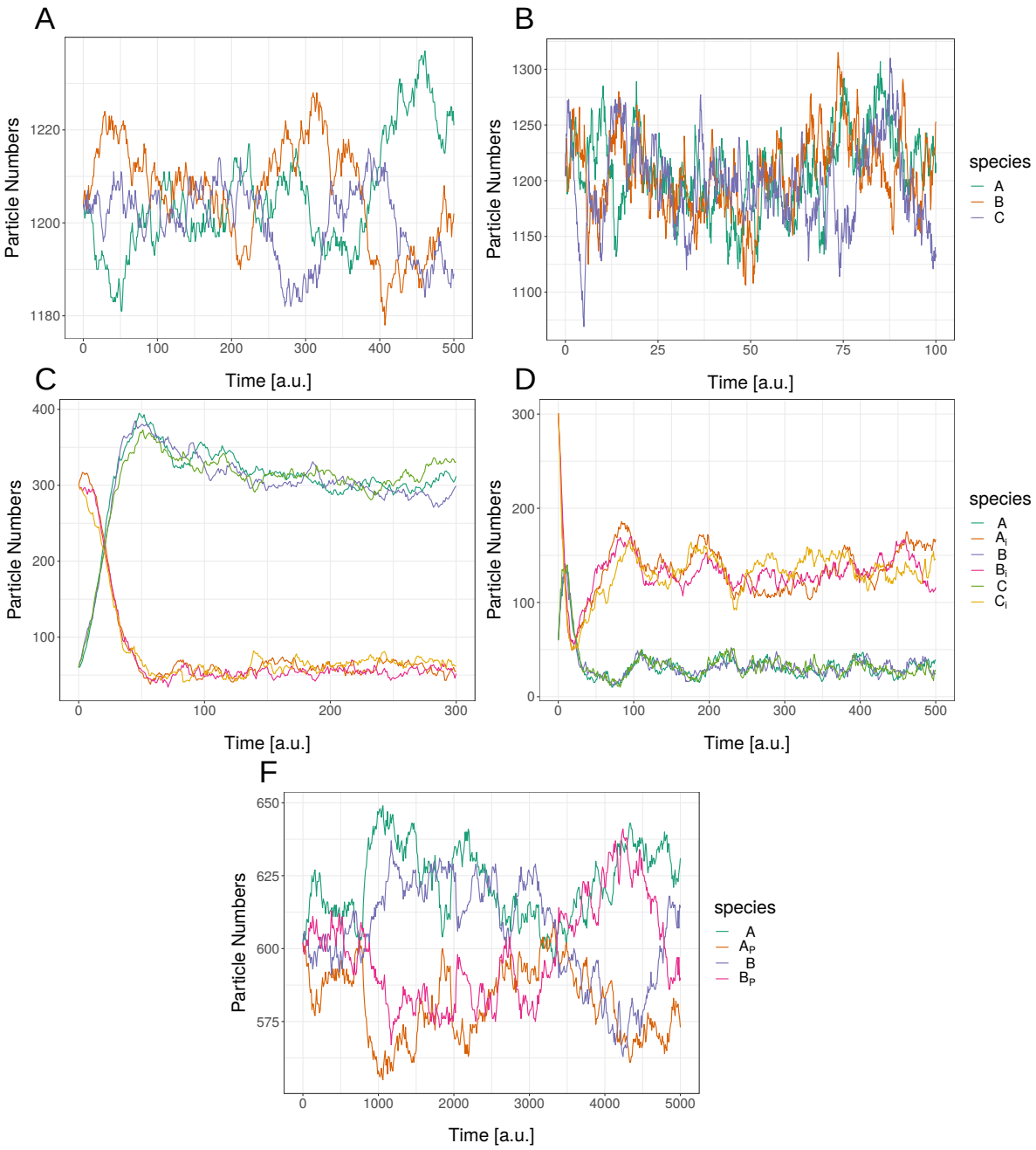

Figure 3: Time courses of the simple biological models. Time series for model SB1 **(A)**, SB1.2 **(B)**, SB2 **(C)**, SB2.2 **(D)** and SM **(E)** were generated using Direct Method as stochastic simulation algorithm with the sampling frequencies 1, 0.1, 1, 1 and 10 [a.u.] respectively. All models reach a steady-state.

case it is used to test for the behavior of TE only upon non-linear dynamics. For this reason, the implausibility of such mass conversions is accepted here in order to keep the number of species as low as possible. Concerning the outcome of estimations for SB1, we cannot determine a range of sampling frequencies meeting the above mentioned criterion for a TE analysis of connectivity (Fig. 4). With increasing environmental history length, CTE increases for all sampling frequencies. The information transfer cannot be

used to discern coupled from uncoupled species, hence TE analysis does not reflect the connectivity in the system. Except for the sampling frequencies 32 and 64 [a.u.], all TE estimations are visually significant. This outcome is not surprising as the model exhibits

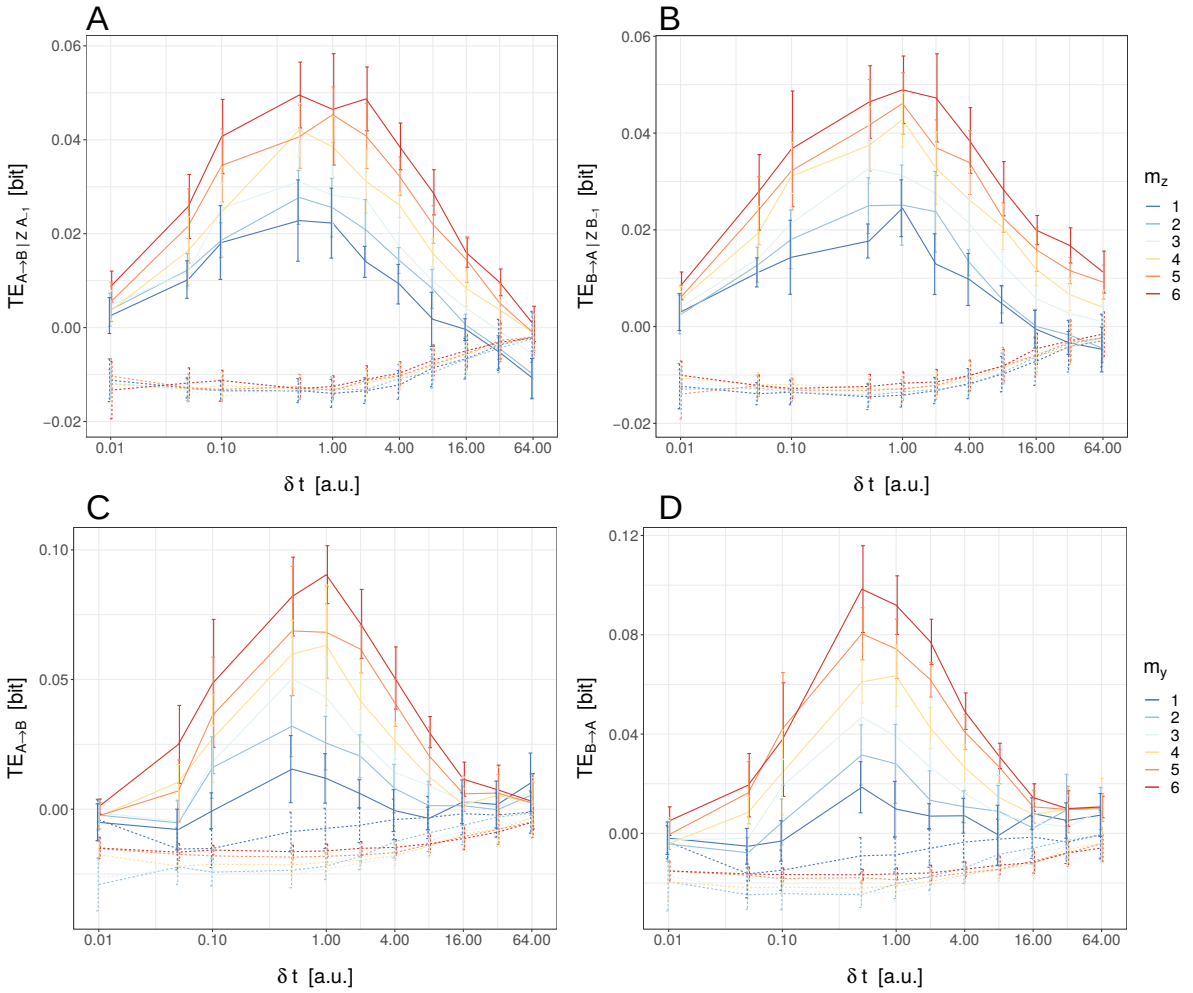

Figure 4: δt-scan for model SB1. TE estimations in dependency of the sampling frequency δt with varying environmental and source history lengths m_z and m_y respectively are shown. Information transfer for different source-target constellations between the species of model SB1 are computed: **A,C:** $A \to B$, **B,D:** $B \to A$. Among these, only $A \to B$ is a truly connected coupling. The upper part displays conditioned, the lower part unconditioned TE. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original or 100 surrogate estimates respectively). Computations are realized with data length $N = 5000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.

high interconnection and complementarity through the mass conversions and the minimal size of three species. Additionally, a mass conversion is difficult to interpret in terms of information transfer. Every change in the substrate concentration finds its complement in the change of product concentration, and the overall mass is conserved. Therefore, in such a constellation, there is no entropy in the information-theoretic sense.

In order to verify whether a model similar to SB1, but constructed as a open system, can be analyzed via TE to infer causality, model SB1.2 was generated. This model is analogous to SB1, but was modified so that every species has its own influx and efflux reactions [ref: model SB1.2, see Appendix dt-scan]. Comparable to the previous results, the TE estimations also do not reflect the underlying structure of the system (see App. A Fig. A.2). But in contrast to SB1, the open system SB1.2 shows a range of sampling frequencies where the information transfer is 0 bits, independent of the history length. These results show that changing the dynamics of the system strongly affects the outcome of TE estimations, even if the underlying connectivity is identical. The idea of model SB1.2 was to add sources of entropy by building an open system, and thus enabling the inference to connectivity by quantifying the information flow in the model. Either the system dynamics are masking the underlying structure, in other words the influx and efflux reactions superpose the effect of the mass conversions, or the general topology of the model and/or the mechanism of mass conversion are unsuitable for such a analysis. In the former case, changing the kinetic parameters could result in a successful TE analysis of connectivity.

In order to build a minimal model which is biologically plausible regarding the connectivity and the kinetic mechanisms, I generated model SB2. SB2 also exhibits a triangular cascade structure, but every species appears in the activated and inactivated form, and the activation process includes another activated species as modifier using the Michaelis-Menten kinetics.
The N-scan of model SB2 shows that the conditioned information transfer of directionally coupled processes does not converge with $N = 32000$, but uncoupled processes reach zero information transfer with > 4000 data points (Fig. 5). In contrast, unconditioned TE converges with > 4000 data points even for coupled processes. This is expected as the unconditioned information transfer is only bivariate and less demanding regarding the data length. We must consider that with the conditioning of TE, the dimensionality of the transition probabilities which have to be estimated in order to calculate information transfer increases with increasing number of environmental processes and increasing embedding dimensions (see Sec. 2.3). Therefore, with increasing dimensionality more data points are required to reach convergent TE and reduce the bias of the estimator. As I am only interested in the qualitative behavior of TE, it is sufficient to use a data length which reflects the qualitative differences between various source-target constellations and the behavior of information transfer upon increasing history lengths. Most importantly, the data length must be sufficiently high to exhibit the convergence of non-coupled processes to zero information transfer, therewith to show the qualitative difference between coupled and non-coupled processes. Hence, I continue to use $N = 5000$ as sufficiently large data

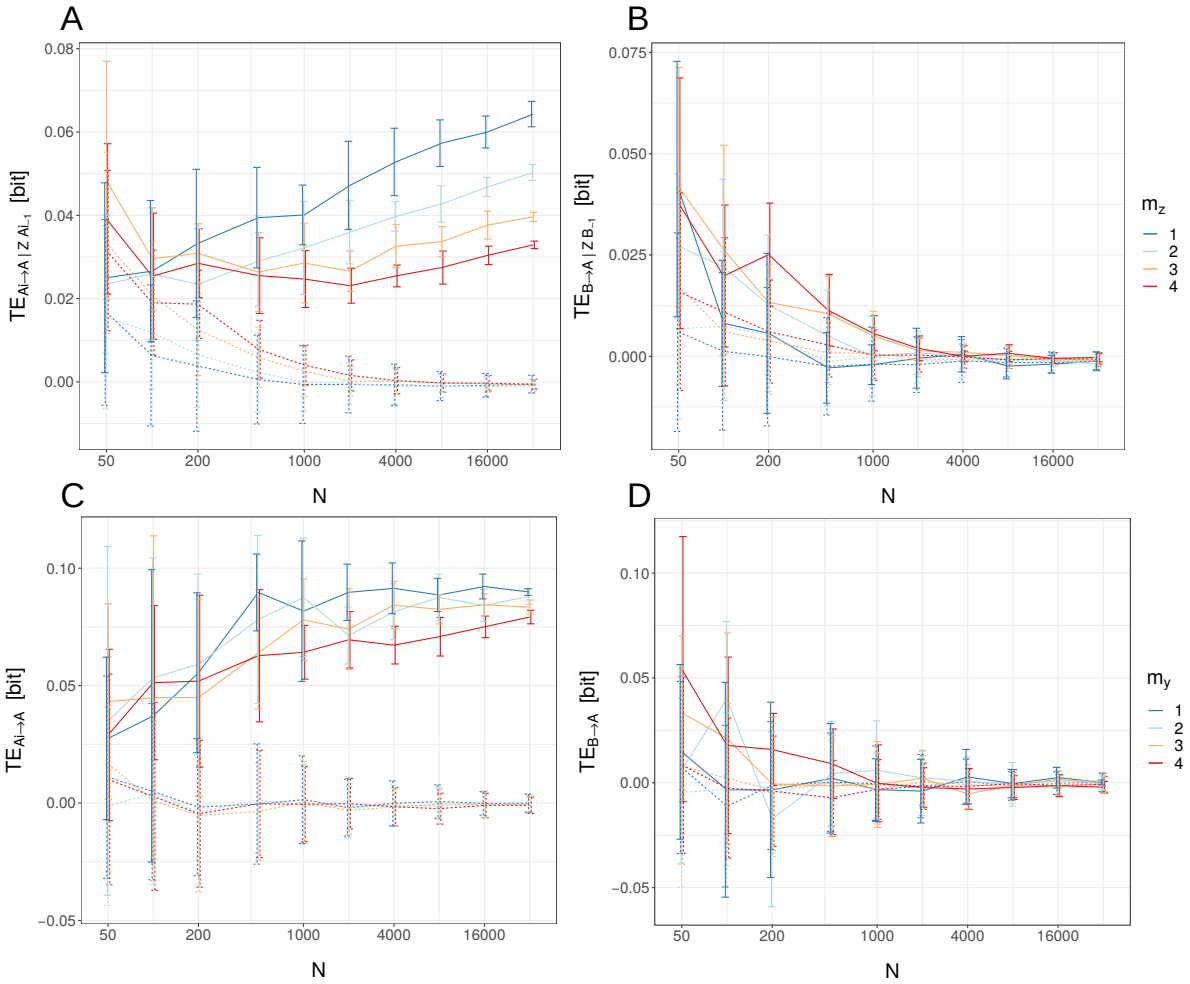

Figure 5: N-scan for model SB2. TE estimations in dependency of the data length N with varying environmental and source history lengths m_z and m_y respectively are shown. Information transfer for different source-target constellations between the species of model SB2 are computed: **A,C:** $A_i \to A$, **B,D:** $B \to A$. Among these, only $A_i \to A$ is a truly connected coupling. The upper part displays conditioned, the lower part unconditioned TE. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (with a total of 10 original and 100 surrogate estimates per data point). Computations are realized with $\delta t = 16$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

length for further TE analysis. As the model size of all (linear and biological) simple models does not exceed the model size of SB2 with 6 species, I conclude that such a data length is also sufficient to reflect the qualitative behavior of TE regarding all simple models.

It is important to mention that for the models SB2 and SB2.2, the significance of the original estimates in difference to the surrogates is calculated as described in Section 2.4 and indicated in the following figures. Note for the δt-scans that only certain sampling frequencies are considered for statistical testing. As the significance tests necessitates higher number of original and surrogate estimations, which increases the computational

effort greatly (see Sec. 2.4), applying these tests to all sampling frequencies was not in the scope of the thesis. Sampling frequencies were chosen for significance testing if these do not exhibit information leak behavior, and thereby are appropriate for further TE analysis of connectivity. As we can see in Figure 6, a range of sampling frequencies can be found where the TE estimation meets the above mentioned criterion and also reflects the underlying structure of the system (δt = [16, 64] [a.u.]). The inactivated species can be identified as significant information sources for the related activated forms. Hence, TE estimation can capture information transfer which is realized as mass conversion between species. Interestingly, those species which act as modifiers for the activation reactions appear as significant sources only for the related inactivated forms, but not for the activated ones (Fig. 7). Explicitly, significant information is transferred in the constellation $C \rightarrow A_i$, but not (or only relatively weakly) for the case of $C \rightarrow A$. Although regarding the kinetics (differential equations and parameters), the modifier acts symmetrically on the activated as well as inactivated species, the information transfer does not seem to reflect this symmetry of connection. Hence, we can observe that the underlying structure is even in the simplest cases of biological, non-linear dynamics not simply reflected in the quantification of information transfer. Nevertheless, no false-positive causal detections are observed for model SB2. Interestingly, this also holds for the unconditioned TE estimations (see App. A Figs. A.3 and A.4).

Subsequently, I changed the system dynamics of SB2 in order to test for the robustness of the TE analysis of connectivity regarding small modifications. For this reason, an efflux reaction for the inactivated species was added, constituting model SB2.2. The δt-scan for SB2.2 shows that the suitable sampling frequency range for TE analysis lies between 8 and 32 [a.u.] (Fig. 8). Here, we can observe that for the activated species A the respective inactivated species and the activated modifying species, A_i and C respectively, can be identified as significant information sources, which is a correct reflection of the connectivity. Additionally, the species C_i and B also appear as significant information sources, although they are not directionally coupled to A. Therefore, the TE estimations infer the underlying structure only partially in a correct way. Considering the structure of model SB2.2, it is important to note that spurious information transfer can be ruled out as through the total interdependency of the system, every process is directionally connected to every other process via a direct or indirect path in the network. Hence in this context, non-causal information transfer is always indirect. Interestingly, the fact which species stands out as significant information source for another process depends on the sampling frequency. Given that the activated species A is the target, for the lower sampling frequency $\delta t = 16$, the activated modifying species C exhibits significant information transfer. But vice versa, for $\delta t = 8$, the respective inactivated species A_i is

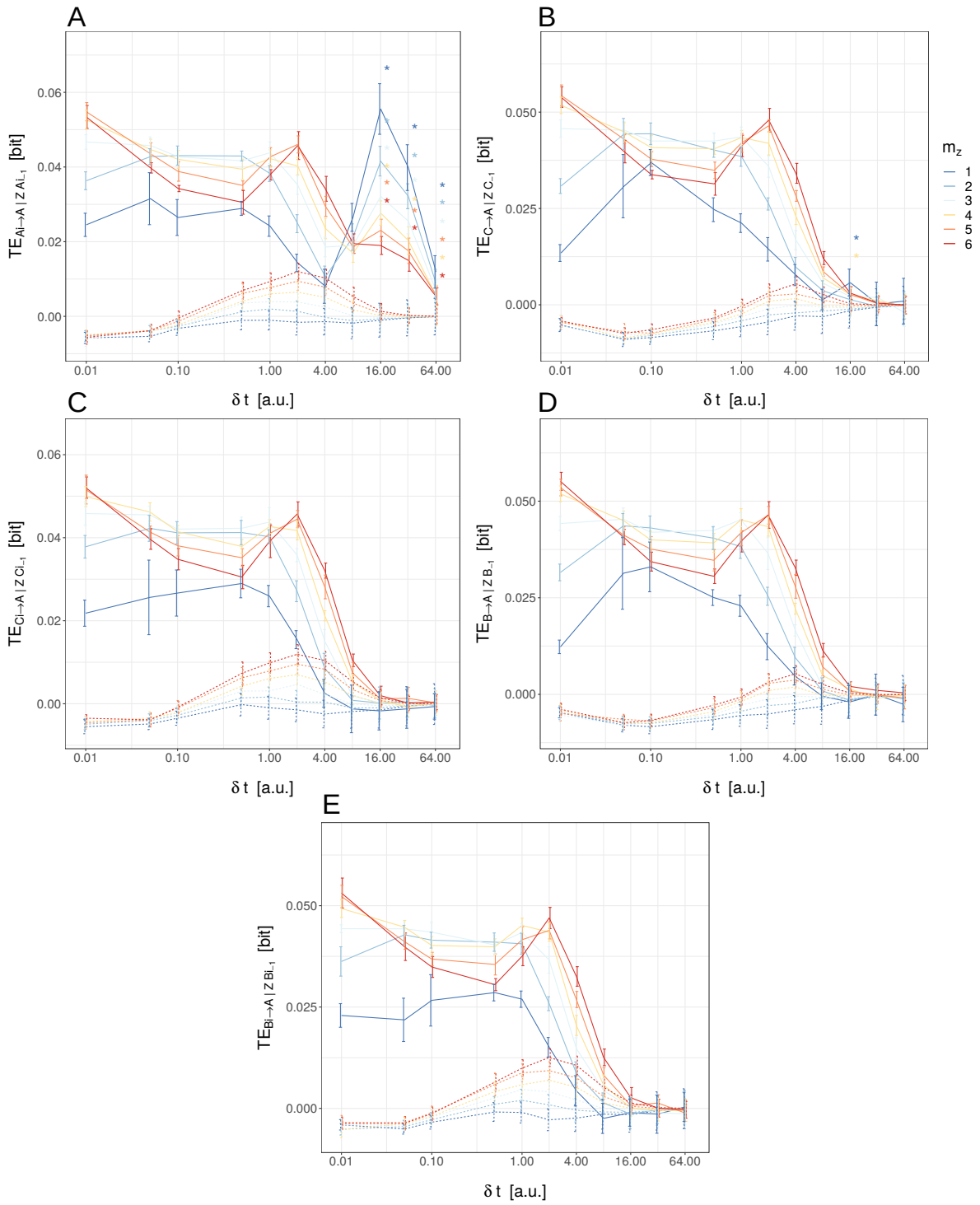

Figure 6: δt-scan for model SB2. Conditioned TE estimations in dependency of the sampling frequency δt and environmental history length m_z for the target species A are shown. Information transfer for different source-target constellations between the species of model SB2 are computed: **A:** $A_i \to A$, **B:** $C \to A$, **C:** $C_i \to A$, **D:** $B \to A$, **E:** $B_i \to A$. Among these, only panel **(A)** and **(B)** display truly connected couplings. Dashed lines indicate the respective surrogate estimations. Asterisks show significance of original estimates with difference to surrogates (one-sample Wilcoxon signed rank test with $p \leqslant 0.01$, Holm-Bonferroni corrected). Note that the significance level is only calculated for δt 16, 32 and 64 [a.u.] with 20 original estimates per data point, where 1000 surrogate estimates correspond to one original estimate. Otherwise, for every original estimate (with a total of 10 original estimates), 10 surrogate estimation were performed, and no significance calculation was applied. Computations are realized with data length $N = 5000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

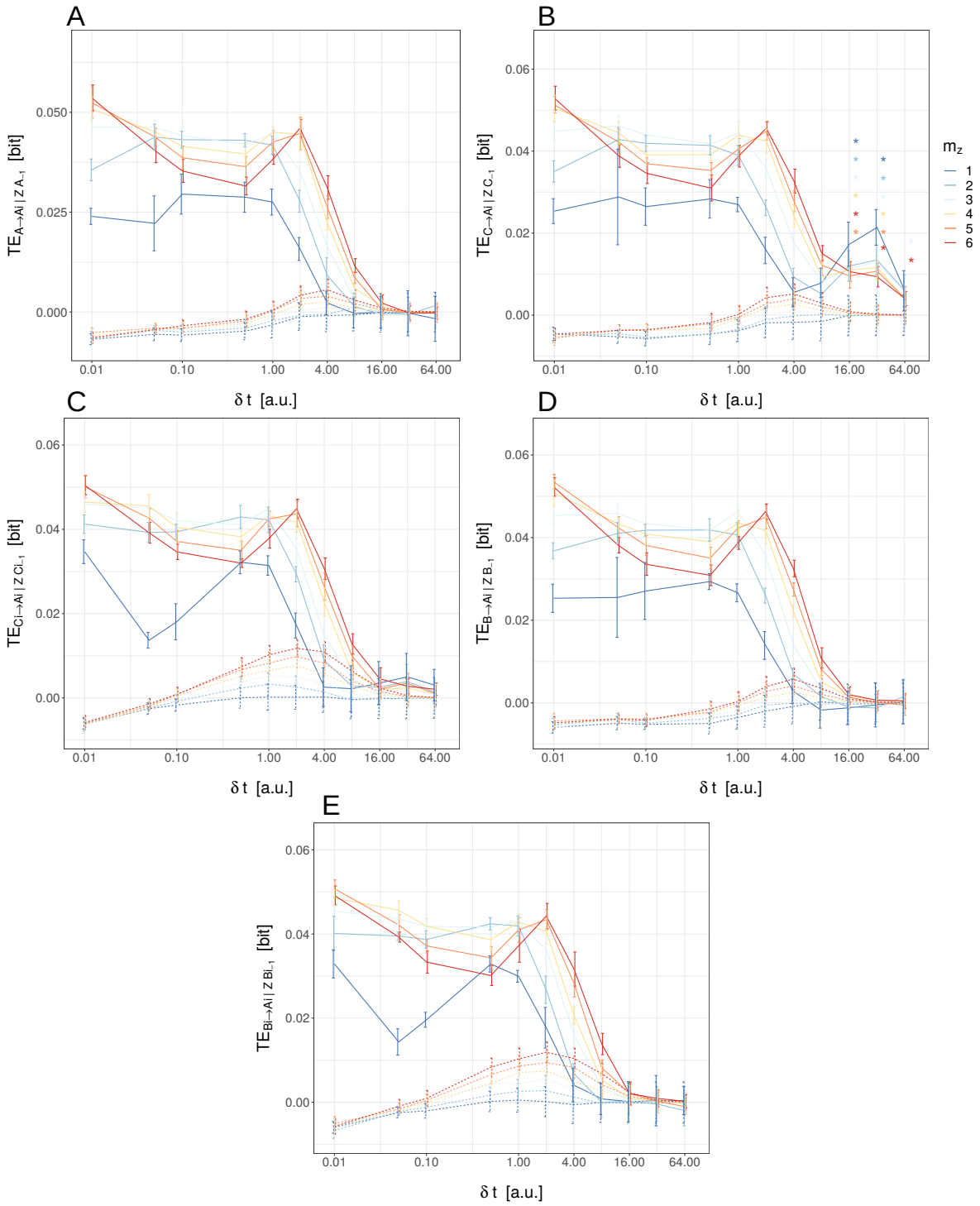

Figure 7: δt-scan for model SB2. Conditioned TE estimations in dependency of the sampling frequency δt and environmental history length m_z for the target species A_i are shown. Information transfer for different source-target constellations between the species of model SB2 are computed: **A:** $A \to A_i$, **B:** $C \to A_i$, **C:** $C_i \to A_i$, **D:** $B \to A_i$, **E:** $B_i \to A_i$. Among these, only panel **(B)** displays a truly connected coupling. Dashed lines indicate the respective surrogate estimations. Asterisks show significance of original estimates with difference to surrogates (one-sample Wilcoxon signed rank test with $p \leqslant 0.01$, Holm-Bonferroni corrected). Note that the significance level is only calculated for δt 16, 32 and 64 [a.u.] with 20 original estimates per data point, where 1000 surrogate estimates correspond to one original estimate. Otherwise, for every original estimate (with a total of 10 original estimates), 10 surrogate estimation were performed, and no significance calculation was applied. Computations are realized with data length $N = 5000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

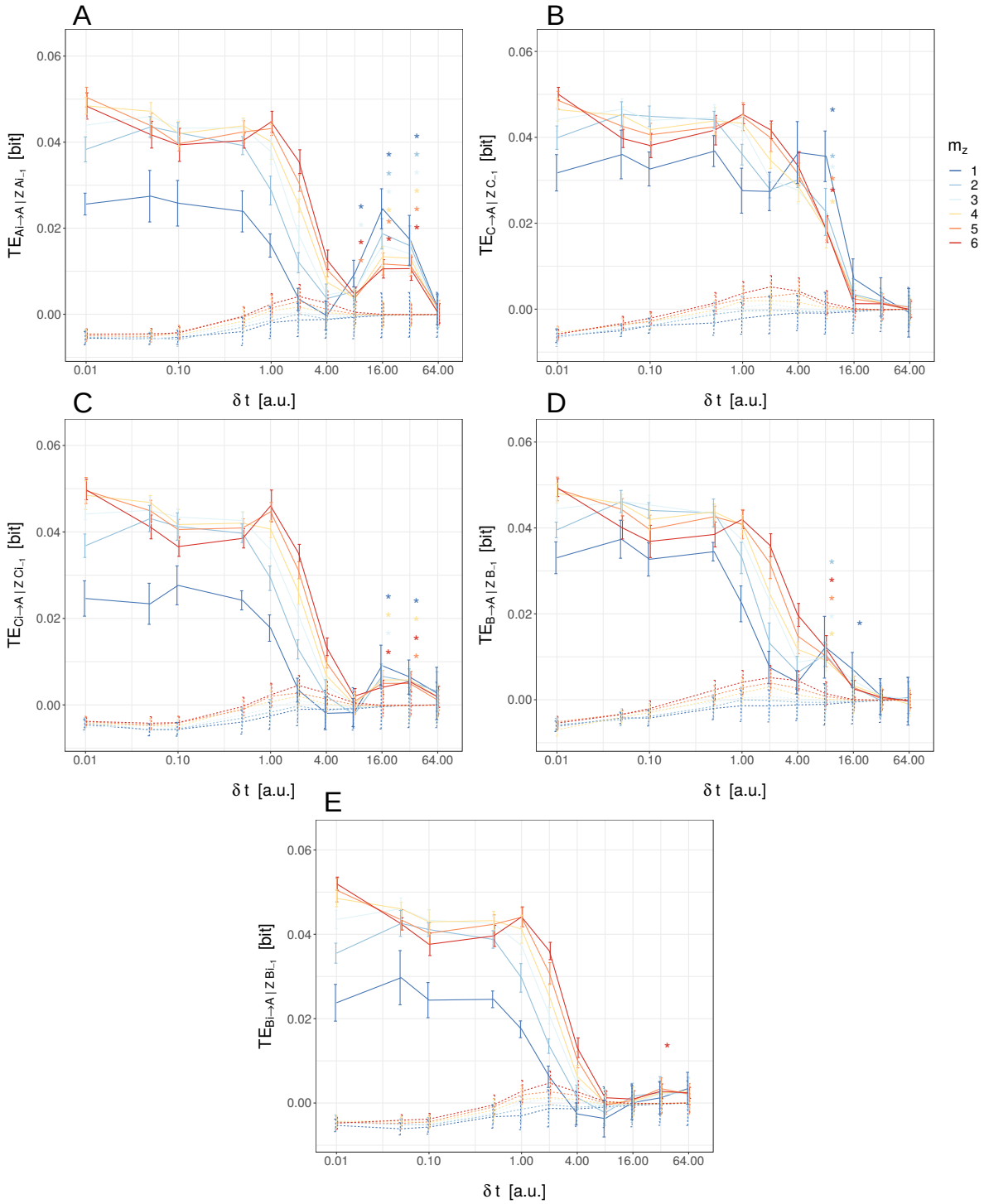

Figure 8: δt-scan for model SB2.2. Conditioned TE estimations in dependency of the sampling frequency δt and environmental history length m_z for the target species A are shown. Information transfer for different source-target constellations between the species of model SB2.2 are computed: **A:** $A_i \to A$, **B:** $C \to A$, **C:** $C_i \to A$, **D:** $B \to A$, **E:** $B_i \to A$. Among these, only panel **(A)** and **(B)** display truly connected couplings. Dashed lines indicate the respective surrogate estimations. Asterisks show significance of original estimates with difference to surrogates (one-sample Wilcoxon signed rank test with $p \leqslant 0.01$, Holm-Bonferroni corrected). Note that the significance level is only calculated for δt 8, 16, 32 and 64 [a.u.] with 20 original estimates per data point, where 1000 surrogate estimates correspond to one original estimate. Otherwise, for every original estimate (with a total of 10 original estimates), 10 surrogate estimation were performed, and no significance calculation was applied. Computations are realized with data length $N = 5000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

the important information source. Hence, the question arises if TE can only quantify the information transfer between species if the used sampling frequency is similar to the underlying causal time scales governing the respective processes.
In Figure 9, looking at the inactivated species A_i as target variable, the most prominent information sources are the activated and as well as inactivated modifying species C and C_i respectively. In this context, the constellation $C_i \rightarrow A_i$ is uncoupled, but exhibits significant information transfer. With increasingly lower sampling frequency ($\delta t = 32$), that constellation shows even the highest information transfer among the source variables. For the sampling frequencies 8 and 16 [a.u.], other weak, but significant indirect information transfer from A, B and B_i to A_i can be observed. These TE estimations meet the criterion of an adequate choice of δt values, but the outcome contradicts the connectivity in the system. This observation also applies to the unconditioned TE estimations for model SB2.2 (see App. A Figs. A.5 and A.6). Even if no information leak behavior is effectively exhibited, indirect information transfer remains detectable for conditioned as well as unconditioned TE. However, this outcome is consistent with the preliminary results, as for the models SL1 and SB2, the unconditioned TE estimations were qualitatively equivalent to the conditioned ones. It is difficult to determine if these false-positive results are caused by numerical problems, e.g. due to the choice of sampling frequency, or some deficiency of the conditioned TE to eliminate indirect information transfer. Finally, we can observe that too low sampling frequencies ($\delta t = 64$) yield generally insignificant TE estimations.
At this point, it is noteworthy to mention that so far, we could observe a coincidence between the range of sampling frequencies not exhibiting information leak behavior, hence meeting the criterion of a TE analysis of connectivity, and the sampling frequencies for which the surrogate estimates display zero mean. This holds for the models SL1, SL1.2, SB2 and SB2.2. For the other models SB1 and SB1.2, where no suitable sampling frequencies could be detected, the negative information transfer of the surrogate estimates approaches zero mean only in case of very low temporal resolution, e.g. $\delta t = 64$ (see Figs. 4 and A.1 in App. A). By trend, such low temporal resolutions induce insignificant and zero information transfer in all previous models, except for the models SB1 and SB1.2, where dependent on the history length, clearly non-zero and visually significant information transfer can be found. These observations can be a hint to general numerical problems of the estimation process specifically related to the models where no reasonable TE analysis could be performed.

Before moving on to the MAPK signaling pathway, let us consider the more minimal model SM which contains the kinetics and motifs of the MAPK model in a simplified context. The goal is to assess the possibility of using information transfer upon structures

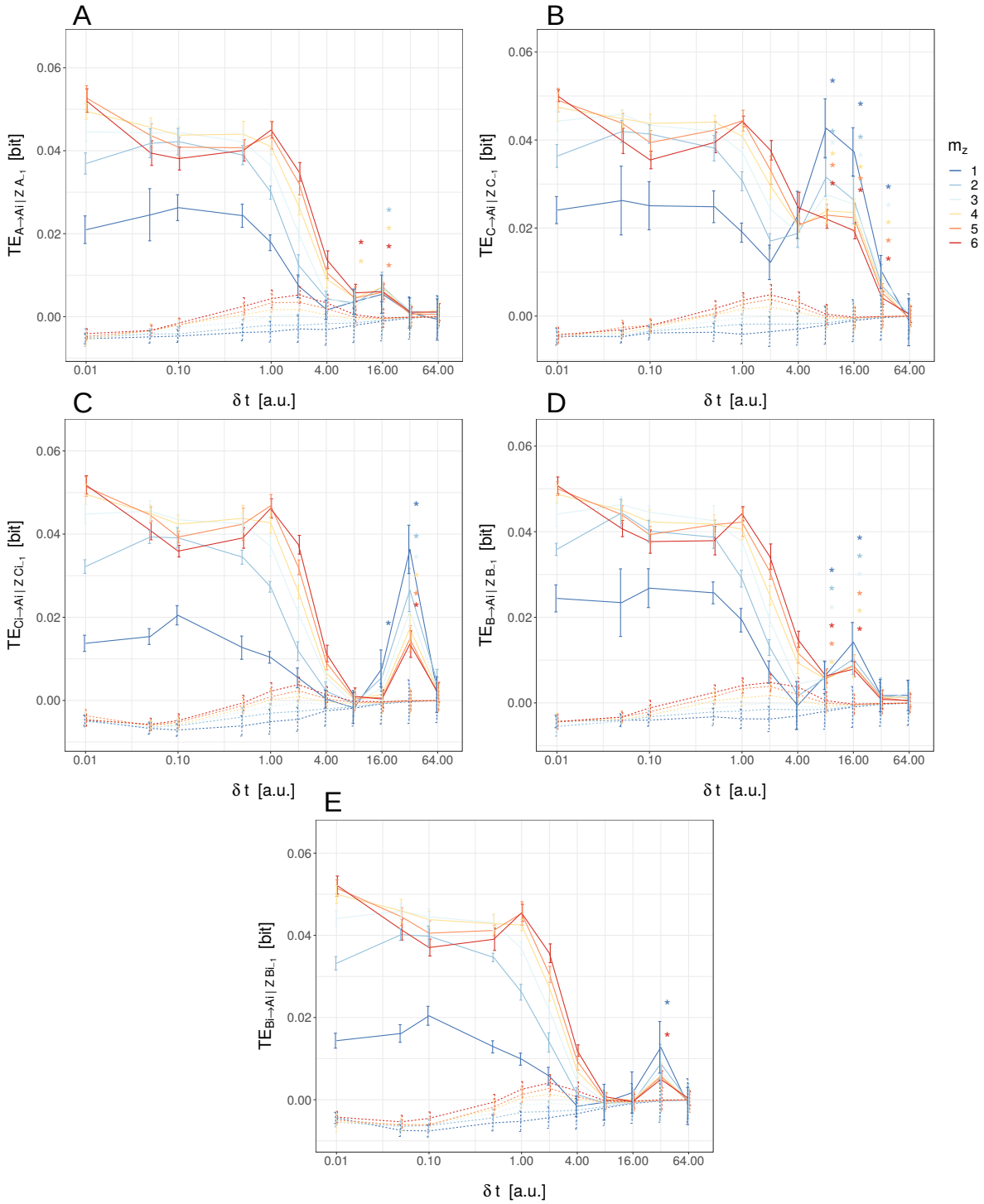

Figure 9: δt-scan for model SB2.2. Conditioned TE estimations in dependency of the sampling frequency δt and environmental history length m_z for the target species A_i are shown. Information transfer for different source-target constellations between the species of model SB2.2 are computed: **A:** $A \to A_i$, **B:** $C \to A_i$, **C:** $C_i \to A_i$, **D:** $B \to A_i$, **E:** $B_i \to A_i$. Among these, only panel **(B)** displays a truly connected coupling. Dashed lines indicate the respective surrogate estimations. Asterisks show significance of original estimates with difference to surrogates (one-sample Wilcoxon signed rank test with $p \leqslant 0.01$, Holm-Bonferroni corrected). Note that the significance level is only calculated for δt 8, 16, 32 and 64 [a.u.] with 20 original estimates per data point, where 1000 surrogate estimates correspond to one original estimate. Otherwise, for every original estimate (with a total of 10 original estimates), 10 surrogate estimation were performed, and no significance calculation was applied. Computations are realized with data length $N = 5000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

which include motifs relevant for biological signal transduction. The kinetics of SM are based on the MAPK model created by Kholodenko, which are all modified Michaelis-Menten kinetics (see Subsec. 2.2.3). SM contains 3 elements which are vital for the MAPK model: A phosphorylation and dephosphorylation cycle of a single species, a phosphorylation cascade (where a phosphorylated species enhances the phosphorylation of another species) and a feedback regulation to an upstream reaction.
Unfortunately, the results of the δt-scan for model SM do not allow us to identify a range of sampling frequencies to continue a TE analysis of connectivity (Fig. 10). All displayed information transfer in that figure refer to truly connected processes in SM. But the only uncoupled constellations $A \rightarrow B$ and $B \rightarrow A$ exhibit identical behavior (data not shown). Independent of the choice of δt, the information transfer increases with increasing environmental history length. Interestingly, the unconditioned information transfer displays irregular behavior (see App. A Fig. A.7). Not only do the original estimates exhibit negative TE values for higher sampling frequencies, but the original estimates are below the surrogate ones for a certain range of temporal resolution. For CTE, this behavior is observable for the sampling frequencies 32 and 64 [a.u.]. For the unconditioned estimations, this irregularity is only observable for the information transfer between the phosphorylated and unphosphorylated form of a moiety (Fig. A.7 C, D). This mechanism is similar to mass conversion which induced problems in the TE estimation before (compare to results of SB1 and SB1.2). But in this case, unconditioned TE produces significant numerical errors. Interestingly, the general outcome of TE estimations is similar to the results of SB1 and SB1.2, where the information transfer analysis also failed to reveal the underlying connectivity. The results show us how difficult it is to perform a TE analysis of connectivity upon highly interdependent networks which exhibit non-linear dynamics, even if all processes are observable can be included in the conditioning of TE. In particular, the outcome of the simplified MAPK model may hint to the fact that analyzing biological motifs which are recurrent in signal transduction networks via information transfer is severely hampered for different reasons, which will be discussed in detail in Chapter 4.

3.2. MAPK Signaling Pathway

Information flow is a main functionality in signaling pathways for which reason the application of information-theoretic measures can give new insights into these biological networks. Different, ubiquitous biological motifs in the structure of the MAPK model, as

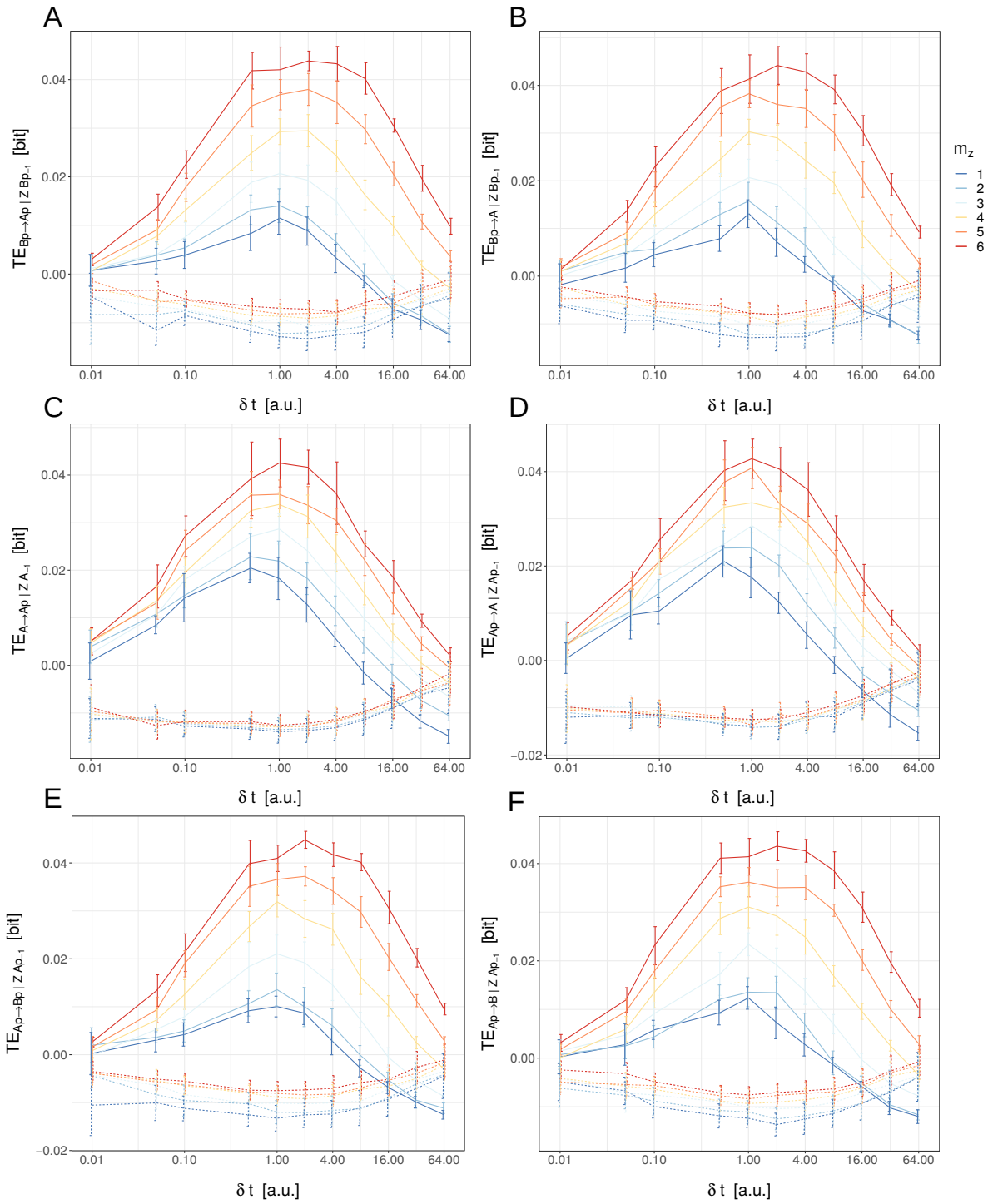

Figure 10: δt-scan for model SM. Conditioned TE estimations in dependency of the sampling frequency δt and environmental history length m_z are shown. Information transfer for different source-target constellations between the species of model SM are computed: **A:** $B_P \to A_P$, **B:** $B_P \to A$, **C:** $A \to A_P$, **D:** $A_P \to A$, **E:** $A_P \to B_P$, **F:** $A_P \to B$. All shown source-target constellations are truly connected couplings. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original or 100 surrogate estimates respectively). Computations are realized with data length $N = 5000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.

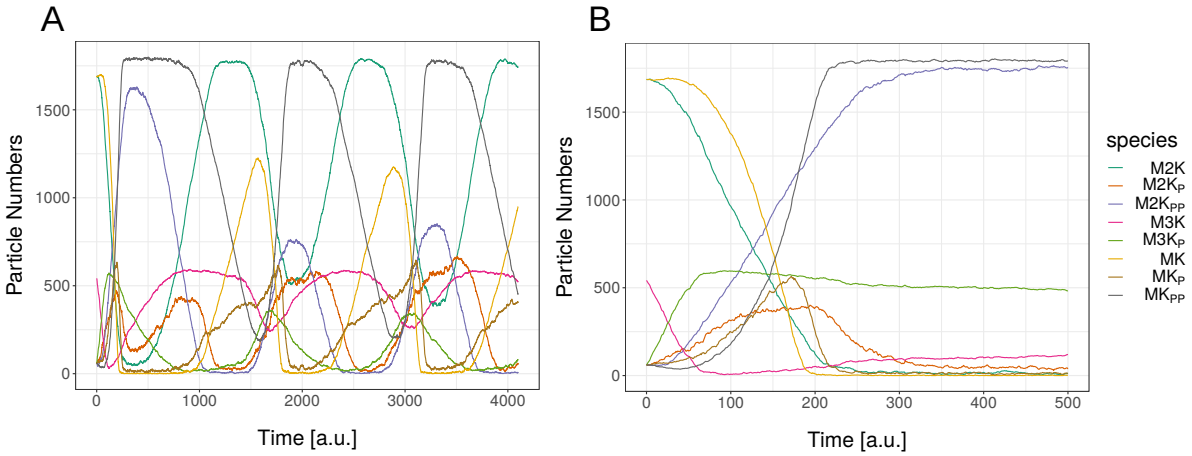

Figure 11: Time courses of the MAPK models. Time series for model M1 **(A)** and M2 **(B)** were generated using Direct Method as stochastic simulation algorithm. The shown instances were simulated with the sampling frequency $\delta t = 1$ [a.u.].

well as the simple Michaelis-Menten kinetics render this system an adequate model for TE analysis of connectivity. TE estimations were performed for two different dynamic states of the MAPK model: for the oscillatory (M1), as well as for the steady-state version (M2) of the system (see Subsec. 2.2.3 and Fig. 11). Although M1 does not exhibit a steady-state, the system can be considered as stationary in the weak-sense [115] as the oscillations are persistent. In other words, the joint probability distributions are preserved if the time series is shifted a certain time interval corresponding to the period of the oscillations, or if the time course is long enough to comprise multiple oscillations so that the exact time point of the end of the time series decreasingly impacts the probability distributions.

In Figure 12, we can see that the N-scan for model M1 shows a peak about $N = 1000$ for conditioned TE, which subsequently decreases and seems to converge to a lower value. Analogously, unconditioned TE exhibits peaking at about 200 data points, followed by a clear convergence. This is unexpected as we would anticipate a similar behavior displayed in Figure 5 for model SB4. Numerical instability at low data lengths could be the reason for the observed outcome. On the basis of these results, it is difficult to conclude if $N = 20,000$ (data length I set for the δt-scans using the MAPK models) is qualitatively sufficient for a TE analysis of connectivity. We cannot observe a qualitative difference between coupled and uncoupled processes as the uncoupled processes do not converge to zero information transfer with the given range of data lengths. Of course, this could be due to the choice of the sampling frequency, which is set to $\delta t = 16$ for the N-scan. It is possible that higher data lengths than $N = 32,000$ are needed to observe a (near) convergence of uncoupled processes to zero information transfer. However, this is computationally not feasible in the context of this work as with $N = 64,000$,

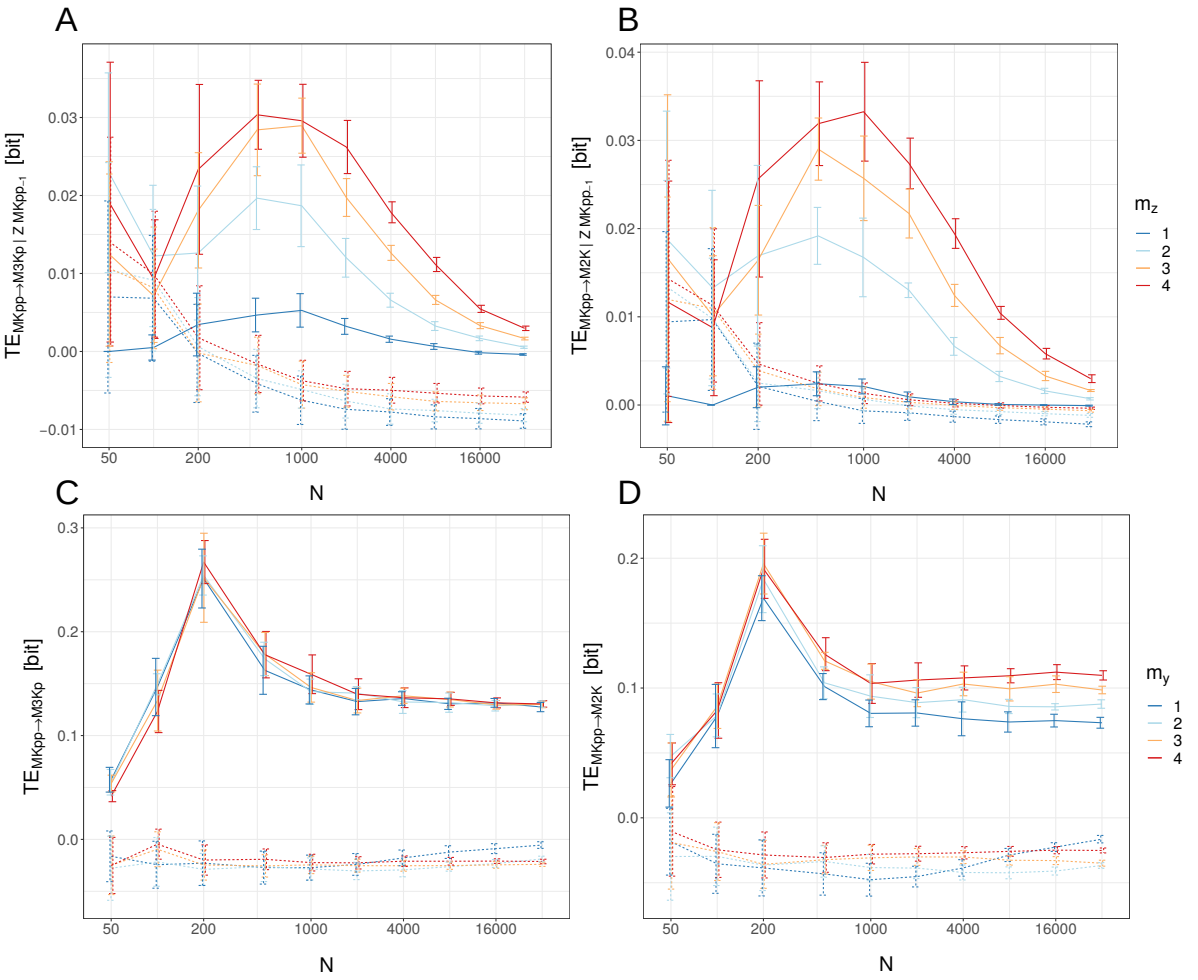

Figure 12: N-scan for model M1. TE estimations in dependency of the data length N with varying environmental and source history lengths m_z and m_y respectively are shown. Information transfer for different source-target constellations between the species of model M1 are computed: **A,C:** $MK_{PP} \rightarrow M3K_P$, **B,D:** $MK_{PP} \rightarrow M2K$. Among these, only $MK_{PP} \rightarrow M3K_P$ is a truly connected coupling. The upper part displays conditioned, the lower part unconditioned TE. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (with a total of 10 original and 100 surrogate estimates per data point). Computations are realized with $\delta t = 16$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

the calculation for a single source-target constellation with a certain parametrization could take several weeks to finish. Possibly, even such high data lengths would not be sufficient to reach convergence, or at least, a qualitative difference between coupled and uncoupled processes. Establishing a whole analysis for different constellations with varying embedding dimensions was therefore not in the scope of this work for higher data lengths.

The results of the δt-scans are shown in Figures 13 and 14 for model M1 and M2 respectively. For both variants of the MAPK model, the results reveal that the identification

of suitable sampling frequencies for TE analysis is difficult, as the criterion that the information transfer should not increase with increasing history length does not hold for all source-target constellations. For the oscillatory model M1, the estimations with the sampling frequencies between 8 and 16 [a.u.] partially meet the criterion of TE analysis, but only in the cases where the information transfer concerns the motif of negative feedback and (de-)phosphorylation cycle (Fig. 13 A-D). But also in that cases, coupled processes cannot be discerned from uncoupled ones, both display positive information transfer. Therefore, the information transfer is not reflecting the underlying connectivity. For the steady-state model M2, the results are more diffuse as most of the original estimates not only exhibit zero information transfer, but also show negative or even smaller values than the surrogate estimates (Fig. 14 E, F). This can be a strong indicator for numerical problems in the estimation process. Hence, these TE estimations cannot be used for further TE analysis of connectivity.

For model M1, the surrogate estimations are mostly negative and approach zero information transfer either for very high or very low temporal resolution, with δt 0.05 or 16 [a.u.] respectively. Again, a similar behavior was observable for the models SB1, SB1.2 and SM for the surrogates before. Concerning model M2, most of the surrogate estimates peak at a certain range of sampling frequencies and subsequently level to zero information transfer. The respective unconditioned TE estimations are shown in Appendix A Figures A.8 and A.9, where we can also see that no discrimination between directionally coupled and uncoupled processes is possible. Here, model M1 either peaks or increases with increasing sampling frequency, whereas M2 levels to zero information transfer with increasing temporal resolution.

Quantitatively, the results for model M1 and M2 differ compared to the simplified MAPK model SM. The information transfer of the MAPK models do not show a bell curve like behavior upon varying sampling frequencies, model M1 exhibits sampling frequencies without information leak behavior, and model M2 levels to zero information transfer. But what is common is the strange behavior of the surrogate estimates which are either positive and (partially) higher than the original estimates (model M1), or persistently negative (model M2 and SM). As the surrogates preserve the probability distribution of the time series, the outcome concerning the MAPK models (but also SB1, SB1.2 and SM) may be a hint for the behavior of the estimator whose bias increases in the face of highly dependent probability distributions. In other words, strongly interconnected networks, as M1 and M2, will induce high statistical dependencies upon the variables in the system. Such statistical dependencies can increase the estimator bias, which will be particularly discussed in Section 4.3. This also may hold for the probability distributions concerning the surrogate estimations. Additionally, if the surrogates do not disrupt all temporal correlations between source and target variable, some dependency of the

variables remains, biasing the surrogate estimations furthermore. Moreover, the surrogate estimates of model M2 increase with increasing history length for some constellations (see Fig. 14 A,B). As the temporal correlations between source and target variable should be disrupted, it is unexpected that changing the history length impacts the surrogate estimates significantly. A detailed discussion about these irregularities and the possible reasons why a successful inference of the underlying connectivity of model M1 and M2 could not be carried out is given in the next chapter.

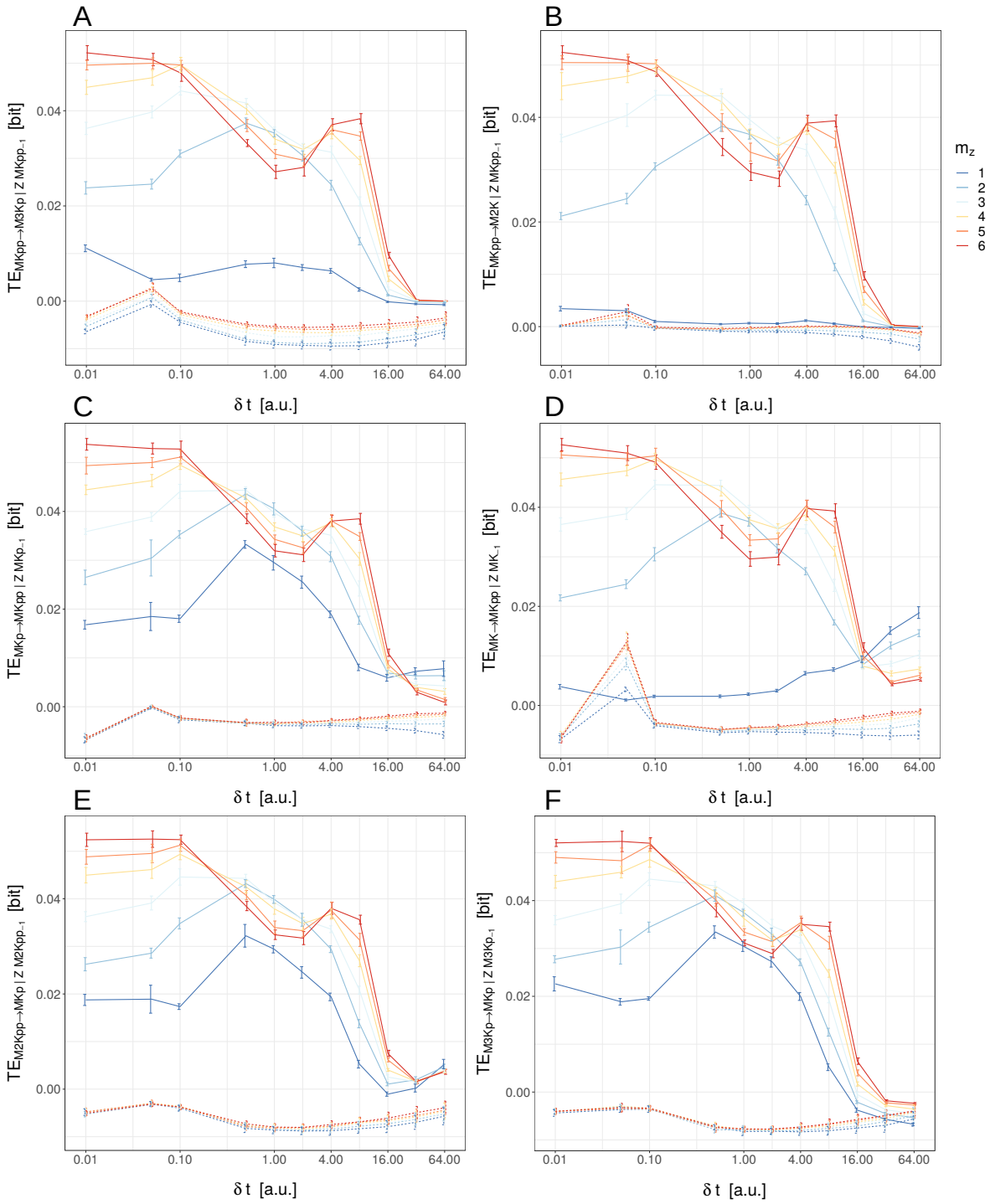

Figure 13: δt-scan for model M1. Conditioned TE estimations in dependency of the sampling frequency δt and environmental history length m_z are shown. Information transfer for different source-target constellations between the species of model M1 are computed: **A:** $MK_{PP} \rightarrow M3K_P$, **B:** $MK_{PP} \rightarrow M2K$, **C:** $MK_P \rightarrow MK_{PP}$, **D:** $M2K_{PP} \rightarrow MK_P$, **E:** $M2K_{PP} \rightarrow MK_P$, **F:** $M3K_P \rightarrow MK_P$. Panels on the left side exhibit truly connected couplings, whereas panels on the right side are unconnected constellations. The chosen source-target constellations correspond to possible information flow via different motifs: **A,B:** negative feedback, **C,D:** (de-)phosphorylation cycle, **E,F:** phosphorylation cascade. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original or 100 surrogate estimates respectively). Computations are realized with data length $N = 20000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.

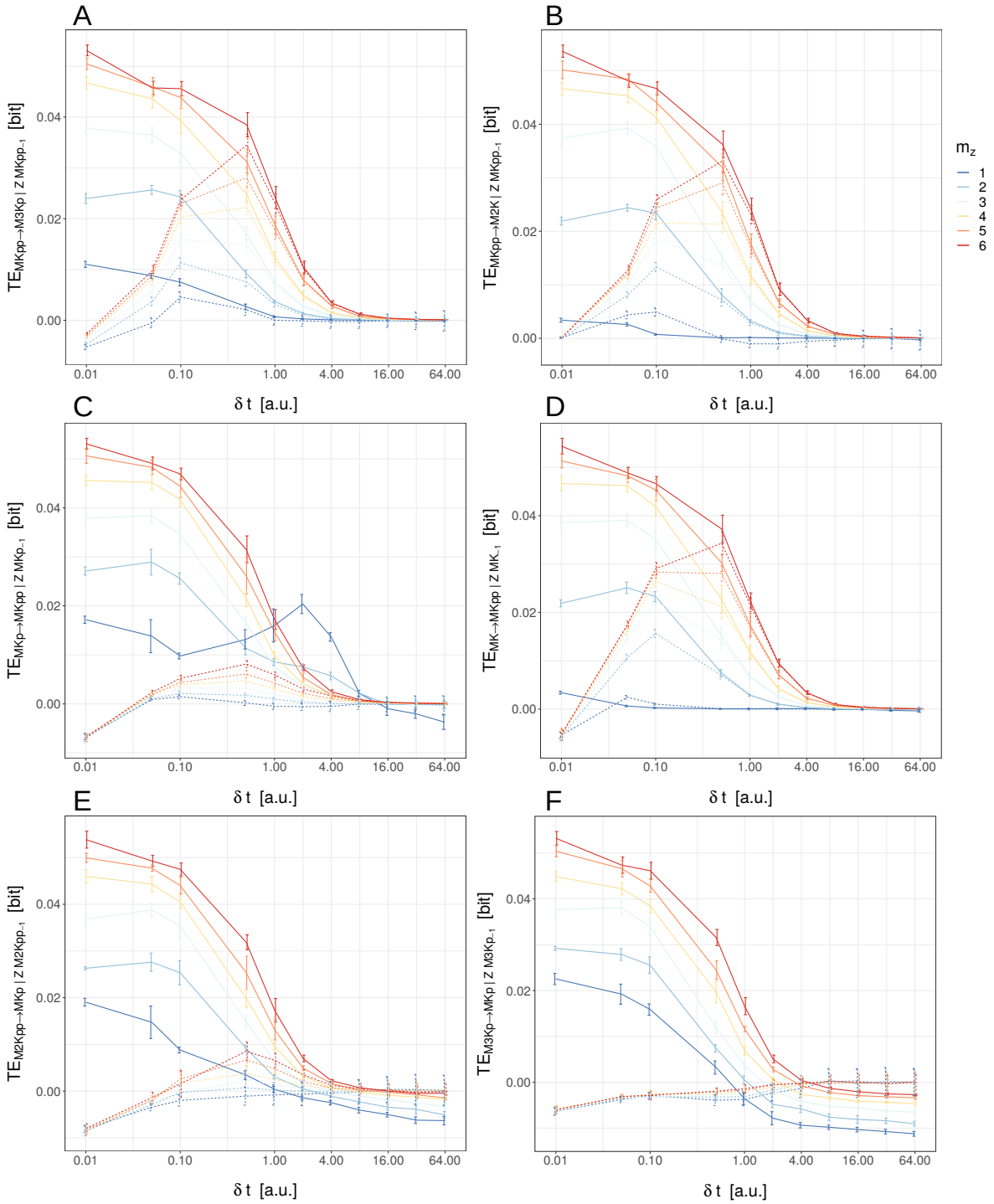

Figure 14: δt-scan for model M2. Conditioned TE estimations in dependency of the sampling frequency δt and environmental history length m_z are shown. Information transfer for different source-target constellations between the species of model M2 are computed: **A:** $MK_{PP} \rightarrow M3K_P$, **B:** $MK_{PP} \rightarrow M2K$, **C:** $MK_P \rightarrow MK_{PP}$, **D:** $M2K_{PP} \rightarrow MK_P$, **E:** $M2K_{PP} \rightarrow MK_P$, **F:** $M3K_P \rightarrow MK_P$. Panels on the left side exhibit truly connected couplings, whereas panels on the right side are unconnected constellations. The chosen source-target constellations correspond to possible information flow via different motifs: **A,B:** negative feedback, **C,D:** (de-)phosphorylation cycle, **E,F:** phosphorylation cascade. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original or 100 surrogate estimates respectively). Computations are realized with data length $N = 20000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.

4. Discussion

Information theory can be a powerful framework to analyze data on the assumption that different, separable entities of the underlying system share information through a common channel. Quantities such as mutual information and transfer entropy are used as model-free approaches for the analysis of different kinds of theoretical and empirical networks for which transferring and processing information is regarded as their main function [48]. In this context, especially transfer entropy is considered a promising candidate to reveal the connectivity in a system through quantifying the information flow using empirical data [116]. One of the most advantageous properties of transfer entropy is that it is dynamic and directed measure to capture information transfer [17]. Nonetheless, the limits and caveats of information transfer analysis have to be considered for every case adequately. Besides the basic requirements of stationarity (or ergodicity) concerning the dynamics of the system and the appropriate parameter settings of a convergent estimator of continuous entropy [58], other limitations concerning the properties of the data (data length, sampling frequency) have to be considered too [55–57, 117].
In the following discussion, I dedicate most of my considerations to synthetic systems. Indications about the deficiencies regarding the application of information transfer analysis to experimental data are mentioned only in some cases. Here, the focus lies on the reflection about the possibilities and caveats of using transfer entropy to infer causality even if all relevant processes in a system are observable.

4.1. System Dynamics, Topology and Transfer Entropy Analysis

The conditioned transfer entropy is an approach to discriminate between direct and indirect information transfer, and therefore enables the possibility to infer the connectivity. Especially in networks with strong interdependency, unconditioned transfer entropy can

quantify high portions of non-causal information transfer, which produces ambiguity in the interpretation of the information-theoretic measure. Conditioned transfer entropy is an approach to overcome this problem, but it can involve other drawbacks. Generally, it is computationally more demanding and requires higher amounts of data. The transfer entropy values decrease with conditioning of environmental processes, in some cases to the extent that significant sources, which can be identified with the unconditioned version, become insignificant with the conditioned information transfer. Above all, if no information about the underlying structure is given, the conditioning has to include all potential information sources, which can strongly decrease the information transfer, or increase the bias of the estimator. These are possible reasons why the transfer entropy analysis upon the MAPK model was not successful for discerning coupled from uncoupled processes. In particular, the results of model M1 (oscillatory MAPK model) could be due to numerical problems linked to estimator bias. As the oscillating behavior generates highly heterogeneous joint and transition probability distributions, which additionally exhibit high dependency on account of the underlying interdependent network, the estimator bias is prone to increase. Details about such bias and the KSG estimator are discussed in Section 4.3. At this point, it is important to mention that dynamics like oscillations showing steep changes in the absolute values can hinder a reliable estimation. Nonetheless, model M2 reaches steady-state, but transfer entropy is still unable to reflect the underlying connectivity. Especially the behavior of the surrogates which partially display higher information transfer than the original estimates hint to numerical instability in the estimation process. Although the dynamics of M2 induce homogeneous probability distributions, the estimator produces considerable bias, even negative values for the original data. Additionally, the effect of generally insignificant information transfer due to total conditioning can be a reason why M2 shows zero information transfer for high temporal resolutions. In general, we can conclude that not only do strongly changing dynamics hamper the estimation, but also the topology of highly interdependent networks which produce comparably dependent probability distributions. The influence of the network structure upon transfer entropy is mentioned in the following sections of the discussion specifically, and presents itself as a recurrent motif of my argumentation. Of course, other conditions as the data length, sampling frequency and embedding dimensions as data and estimation dependent parameters have to be taken into account and are discussed below.

4.2. Data Length and Transfer Entropy

Regarding the dependency of information transfer on the data length, we saw that the range where transfer entropy converges depends on the examined source-target constellation. Although the herein used data lengths (5000 and 20000 data points for the simple and the MAPK models respectively) do not lie in the range of convergence for all constellations, the data length of the simple models is supposed to be sufficient to reflect the qualitative behavior of transfer entropy which enables causal inference. In other words, we could observe the qualitative difference between various source-target constellations, the behavior upon variant data lengths and embedding dimensions, and especially zero information transfer for directionally uncoupled processes. For that last point, it is not necessary to reach convergent transfer entropy in dependency of the data length for coupled constellations. Convergence to zero information transfer was achieved with fewer data points. Hence, it is not to expect that a further increase of data length will qualitatively change the presented outcomes regarding the simple models. In contrast, it is possible that the MAPK models need substantially higher data lengths to perform an information transfer analysis which could reveal the underlying connectivity. Nonetheless, it is questionable that the increased model size of the MAPK models alone (8 species in total, 2 additional species compared to SB2) would result in such a massively increased demand for data points. As mentioned above, it is more likely that the specific dynamics of the MAPK models, statistically speaking the probability distributions, posed a problem for the here constructed estimation process. Even for the simpler variant of model SM, which consists of 4 species, the information transfer analysis failed to reflect the underlying causal structure. Therefore, I think that the reasons for the observed deviation from correct causal inference lie beyond the insufficiency of the data length.

4.3. Estimator Bias

Concerning statistical errors in the estimation process, the KSG algorithm is reported to be fairly stable [49, 118, 119], and therefore the statistical bias is considered as marginal compared to other effects caused by the embedding dimensions, conditioning or structure of data. Nonetheless, the work of Gao et al. [66] shows that for strongly dependent variables, the "KSG estimator tends to systematically underestimate mutual information", as summarized by Singh et al. [67]. The tendency of underestimation was also reported for transfer entropy by Surovtsova et al. Regarding the highly interdependent structures

of the models analyzed in this thesis, the estimator bias for dependent variables may strongly limit the analysis of biological examples and induce high rates of false-negative causal detections. Signal transduction networks are strongly interconnected and can be particularly difficult to analyze via transfer entropy using the KSG algorithm. Especially feedback loop motifs generate high statistical dependencies between the variables. Such statistical dependencies increase the non-uniformity of the probability distribution estimated via the KSG method [66] – more precisely, the joint and transition probability distributions are meant, as it can be seen from Equation (3). Additionally, if partial or total conditioning of environmental processes is performed for the transfer entropy estimation, the dimensionality of the probability distributions increases, and with it the estimator bias. More complex dynamics than steady-state, e.g. oscillations, will exhibit more heterogeneous probability distributions, hence even more non-uniformity and estimator bias is expected. For this reasons, one should consider to use an estimator which attenuates the local uniformity assumption which is adopted by the KSG algorithm [64]. Gao et al. developed an estimator which corrects for non-uniformity by using principle component analysis (PCA). The performance of this improved k-nearest neighbor estimator can be used for future projects trying to shed light on biological networks with more complex dynamics.

We could observe the coincidence between the surrogate estimates distributed around 0 bits and information transfer reflecting the model structure. It can be argued that this finding should not be considered as necessary criterion for applying a transfer entropy analysis for causal inference. As mentioned before, the KSG estimator is based on the k-nearest neighbor statistics, which is not a kernel density estimation algorithm. Therefore, negative estimates are (also theoretically) possible and not necessarily a hint for numerical instability of the estimation process. Nevertheless, it is an interesting question to investigate what the behavior of transfer entropy upon surrogate data indicates about the information flow in a system.
More interestingly in this context is the consistency of the surrogate estimate behavior across different models. All models for which transfer entropy could not reveal the underlying connectivity (see results of models SB1, SB1.2, SM, M1 and M2) exhibit surrogate estimations which are either persistently negative, or positive and above the original estimates. This outcome can be interpreted as a strong indicator of numerical instability in the estimation process. The surrogate data preserves the probability distribution of the original time course, but temporal correlations of the information source should be destroyed. Hence the specifically temporal contribution of the source variable to the transition probabilities of the target is expected to be disrupted. However, the possibility remains that the statistical dependencies between the variables are

conserved throughout surrogate data. In other words, as the probability distribution of the source variable is preserved, the transition probabilities of the surrogates may share some similarities with the transition probabilities of the original data. Therefore, the surrogate estimates do not only serve to correct for the bias of the estimated transfer entropy values, but can serve as indicators of the estimator bias in face of the given probability distributions. The former bias is due to the structure of the data and expected to be revealed by creating surrogates (we can refer to it as estimation bias), whereas the latter bias is only a result of the estimator, labelled as estimator bias. Hence, estimation and estimator bias should be distinguished. Additionally, if the temporal correlations between the processes are disrupted via surrogate data, then altering the history lengths should not impact the surrogate estimates significantly. Nonetheless, this does not hold for model M2 where some temporal correlations between source and target seem to be conserved. Such an outcome undermines the validity of surrogate data in their function as hypothesis test for transfer entropy estimations. From this perspective, I suggest to use the surrogate estimations as indication about the estimator bias, especially if the surrogate estimates are higher than the original ones, or if these exhibit negative values. Thereby, the surrogates can help to identify appropriate parameters, e.g. sampling frequency, for information transfer analysis by excluding the parameters which increase the estimator bias. This is consistent with the outcomes of the models SL1, SL1.2, SB2 and SB2.2 which showed that the surrogate estimates indicating increased estimator bias also coincide with the sampling frequencies which exhibit information leak behavior. In Sections 4.4 and 4.5, other criteria for identifying appropriate parameters, namely the sampling frequency and embedding dimensions, in the context of transfer entropy analysis are discussed. Additionally, the validity of surrogate data, which depends on the capacity to disrupt temporal relations, should be assessed by observing the behavior of the surrogate estimates upon varying embedding dimensions.

4.4. Information Transfer and Sampling Frequency

It should be mentioned that reasonable transfer entropy estimations could be achieved beyond the applied range of sampling frequencies if δt-scan is performed in a more fine-grained or extended range. For the sake of fast computation and conceptual prove, I considered a doubling of the sampling frequency in the range of 0.01 to 64 [a.u.] as sufficient and reasonable. Especially for the MAPK model, such a more fine-grained scan for suitable sampling frequencies could be applied and render successful results. Due to

limitations in time, such testing was not realized and could be performed in the future. More importantly, we could observe that only a constrained range of sampling frequencies were suitable for an information transfer analysis which allows inference to causality. Explicitly, too high sampling frequencies did not satisfy such an adequate analysis. On the other side, too low sampling frequencies can decrease or even yield generally insignificant transfer entropy estimations, increasing the probability of false-negative causal detections. In general, decreasing the temporal resolution of the data can destroy the system dynamics, breaking the causal relationships between variables by distorting the transition probability distributions. Especially by deleting the past of the source variable, the most informative parts for the prediction of the target process can be lost, disabling the detection of information transfer. For illustration, assume a discrete system with the dependent processes X and Y, where the state y_n of Y at time point n affects the transition probability $x_n \rightarrow x_{n+1}$ of X – a simple bivariate system of Markov order 1. If we reduce the temporal resolution of both processes, i.e halving the sampling frequency, then spurious effects are introduced as now the contribution of y_n to $x_n \rightarrow x_{n+2}$ will be quantified. For continuous processes, it is more likely that the transition probabilities can be constructed as $x_n \rightarrow x_{n+k}$, where k is not necessarily a distinct continuous time value, but subject to a probability distribution dependent on the dynamics of the system. Similarly, not only y_n, but also y_{n+l} will contribute to the transition probability of $x_n \rightarrow x_{n+k}$ (again with l subject to a certain probability distribution). From this we can conclude that reducing the temporal resolution of the data beyond possible values of k or l can disrupt the causal relationship between process X and Y. This tendency was theoretically and empirically confirmed by previous work [120–122]. Surprisingly, the results of the δt-scans seem to partially contradict these insights as a lower temporal resolution of the data does not correspond to a decrease of transfer entropy, and can even improve the detectability of coupled processes. Similar results have been reported by past research [56]. Barnett et al. show for the Granger causality that downsampling may potentially increase the quantity and the identifiability of causal links, as a convergence between the sampling frequency and the underlying causal time scales of the system can positively contribute to the detection of connectivity [117]. Furthermore, both concepts, Granger causality and transfer entropy, are related and can be interpreted as measures for directed information transfer [48, 123]. Therefore, my observations concerning the behavior of transfer entropy in dependency of the sampling frequency may be due to a similar relation between the underlying causal time scales and the temporal resolution of the data. Similar to the discrete example above, let us consider that we introduce a time lag τ into the causal relationship from Y to X so that y_n now contributes to the transition probability $x_n \rightarrow x_{n+\tau+1}$. If transfer entropy is calculated without explicitly introducing a time lag in the estimation process, then again the effect of y_n upon $x_n \rightarrow x_{n+1}$ will be

considered. However, the reduction of the sampling frequency can lead to a situation where we consider the true underlying causal relationship between y_n and $x_n \rightarrow x_{n+\tau+1}$.

It is noteworthy to mention that the wrong choice of sampling frequency cannot only render coupled processes undetectable, but also can produce false-positive results [117]. I argue that the portion of direct and indirect information transfer also depends on the temporal resolution. Generally, a higher temporal resolution provides more information to predict the "future" states of a target variable given other processes. This is confirmed by the δ-scans where higher sampling frequencies exhibit higher information transfer, which also accords with past research [55]. Considering information leak, the choice of the sampling frequency could affect the direct and indirect information transfer differentially. As mentioned in Subsection 1.1.2, I define information leak as the observable behavior of increasing transfer entropy upon increasing history lengths. Such behavior can hint to another process in the system whose contribution to the information transfer between a source and a target is not conditioned, hence unblocked and non-causal. Especially if the higher temporal resolution strongly differs from the underlying causal time scale of the processes in question, higher sampling frequencies can decrease the portion of direct information transfer, as mentioned above. Simultaneously, higher sampling frequencies may also amplify existing non-causal information transfer, if the temporal resolution exhibits strong convergence with the underlying non-causal time scales of the environmental processes. On the other side, the sampling frequencies in the transfer entropy analysis may be used to reveal the causal and non-causal time scales governing the processes in question. With regard to biological networks, such an approach can help to uncover the temporal rate of information transfer between specific processes in a system, and between different systems. Particularly for biological structures where temporal dependencies are crucial, as in neuronal networks or circadian rhythms, analyzing the underlying time scales via information transfer can be a promising technique. But also in the context of signal transduction, temporal correlations can be crucial to establish certain dynamics as ultrasensitivity or bistability, as it was reported for the MAPK pathway [124]. From this perspective, transfer entropy analysis can help to clarify the biological functionality of certain structures which generate the governing time scales.

Interestingly, previous research using a rectangular kernel density estimator for transfer entropy quantification showed that the information transfer decreases with increasing sampling frequency [57]. In my case, the information transfer remains (more or less stably) on a relatively high level for high temporal resolutions. Additionally, in this domain of high sampling frequency, transfer entropy is highly sensitive to the history lengths and increases with the embedding dimensions. Besides the possibility of information leak,

both observations concerning high temporal resolutions could result from the behavior of the KSG estimator dealing with highly resolved data. In this regard, numerical instabilities related to the k-nearest neighbor statistics have to be analyzed with respect to the density of the given data.
In a specific domain of sampling frequencies, the estimated information transfer could (partially) reveal the connectivity in the system, depending on the model. While the connectivity of model SB2 was inferred correctly, model SB2.2 displayed false-positive and false-negative causal detections. It was observable that the sampling frequencies (partially) reflecting the underlying connectivity coincided with the transfer entropy values not exhibiting information leak behavior. Therefore, I decided to take this empirical finding as necessary criterion to detect suitable sampling frequencies for an information transfer analysis of connectivity. As reported by Surovtsova et al., if the information transfer increases with increasing embedding dimensions, this indicates either an information leak caused by another important, but unconditioned process in the system, or the wrong choice of the history lengths [69]. It is important to mention that generally, such behavior can also be a result of estimator bias. Although an identical conditioning and parametrization of the transfer entropy estimation was used, the behavior of information transfer upon varying embedding dimensions differed depending on the sampling frequency. In this regard, I suggest to exclude sampling frequencies from further information transfer analysis if information leak behavior is observed. Especially, if the behavior of information leak is not observable for other sampling frequencies under identical conditions.

4.5. Criteria for Optimal Parameters in Transfer Entropy Analysis

Considering the literature, other criteria to choose optimal parameters for information transfer analysis have been presented. Ragwitz et al. propose a locally constant predictor which is a criterion for the reliability of the predictions using the variance of the transition probabilities [74]. Garland et al. introduce a measure called the "time delayed active information storage" which serves to maximize the shared information between the present and the future state of the system [125]. The proposed measures are useful to identify the optimal parameters (time delay and the embedding dimension) for the predictive information transfer. Hence, it would be interesting to check if the sampling frequencies identified in my presented information transfer analysis could also be detected

via the above mentioned criteria in the literature. Moreover, these criteria could also be used to identify an optimal setting of the embedding dimensions. Due to the limited amount of time, such procedures could not be implemented in my analysis and should be applied in prospective projects.

But at this point, it is useful to discuss the differences and convergence concerning the concepts of prediction and causality. The criteria mentioned above are useful in terms of prediction, in other words to maximize the statistical dependence contained in the of time series which are reconstructed according to the embedding dimensions and/or time delay parameters. In contrast, this work aims to statistically infer the connectivity of a (biological) network using predictive information transfer. These are two distinct concepts for which we can compare the two related measures transfer entropy and Granger causality. As mentioned above, both concepts share an inherently statistical nature constructed to identify the predictive capacity of one process over another, and transfer entropy can be regarded "as a non-parametric test statistic for pure Granger causality" [48]. Whether or not such a statistical notion of causality is sufficient to grasp the underlying true causal connections in a system is vividly discussed in the literature [75, 126–130]. Besides a fundamental critique of the notion of causality, the applicability of these concepts to multivariate time series in order to infer Granger causality has to be considered. In this context, we have to note that for non-Gaussian processes, transfer entropy and Granger causality are not identical quantities [48, 53]. Hence, both approaches may converge to similar results regarding the optimal parameters. But first, this convergence is not unconditional, as the quantities are not necessarily identical. Second, even if transfer entropy converges to Granger causality, the optimal set of embedding dimensions to block all non-causal information transfer is not necessarily the same set of parameters which maximize the statistical dependence between a target and source variable. Or in other words, there is no reason to suppose that increased statistical dependence between variables converges to the principle of direct information transfer.

Following this argumentation, finding the optimal parameters, e.g. embedding dimensions and temporal resolution, using the above mentioned criteria in the literature can lead to a situation where one ends up by maximizing the reliability of predictions, while diverging from the goal of detecting the connectivity in a system. Of course, such possibilities have to be confirmed by empirical findings. However, the empirical criterion of using the sampling frequencies which do not exhibit information leak behavior, while other temporal resolutions do, should be included in information transfer analysis. For future projects, this criterion can be combined with other criteria discussed in the past. At the end of the discussion, I outline such an analysis in Section 4.8.

As finding the optimal parameters to infer causality can be highly demanding or practically unfeasible, let us deliberate such a process proposed in the literature in order to identify suitable embedding dimensions. We assume that we have found an optimal set of history lengths which maximize the above mentioned criteria for reliable predictions. We further assume that this optimal set includes the underlying Markov orders of the system, or at least, that it includes the set of history lengths to practically eliminate non-causal information transfer. Referring to the research of Garland et al., these optima can be broad surfaces [125]. Therefore, some heuristics are needed to translate these results in a distinct choice of embedding dimensions. Considering the maximization of reliable predictions, these heuristics will try to account for the trade-off between efficiency and accuracy. Therefore, taking the lowest possible values of the optimal set is advisable as higher history lengths are computationally more demanding. But regarding causality, such minimal heuristics cannot be naively adopted. Then, the best way to practically exclude non-causal information transfer would be to take the highest possible values of the optima (maximal heuristics). But with it, the computational cost and required data length can strongly increase [125]. Especially in large and complex biological networks, adopting such maximal parameter sets can be unpractical. Additionally, if the surface of the optima is broad, the highest optimal set of history lengths can reach dimension which are not manageable. Therefore, in the context of the mentioned criteria for maximizing predictions, simple minimal or maximal heuristics in order to determine distinct parameter sets may not be easily applicable for the goal of inferring causality.
An additional complicating factor is that we have at least three different embedding dimensions for the conditioned transfer entropy (source, target, environmental processes), resulting in a three-dimensional surface of the optima. If we assume high heterogeneity for the underlying Markov orders regarding the environmental processes, we either end up taking the highest optimal value for all environmental processes, or we decide to set different history lengths for these processes. In the first approach, environmental processes with lower underlying Markov orders are reconstructed with a much higher history length. This may result in the increased possibility of rendering significant information transfer insignificant. This general tendency that transfer entropy decreases with increasing history length was observed for VAR processes [69] as well as for the models SB2 and SB2.2. In the second approach of setting different history lengths for environmental processes, it becomes more difficult to apply some heuristics to an n-dimensional surface of optima.
To sum up, an optimal set of embedding dimensions should fulfill the following requirements: high enough to eliminate non-causal information transfer, low enough to guarantee efficiency, differential enough (concerning environmental processes) to prevent an increase of false-negative causal detections. Ultimately, taking account of all the

mentioned caveats of using transfer entropy analysis to infer causality, we have to concede that there can be no practical guarantees to eliminate non-causal information transfer by using some optimal parameter sets. As a general conclusion, I would state that non-causal information transfer remains a part of the transfer entropy analysis.

Regarding the transfer entropy analysis I performed, the δt-scans for the conditioned transfer entropy were realized not only with the variation of the environmental history length, but also by scanning the source and target past lengths. As these variants of transfer entropy estimation did not qualitatively differ from the data which is shown, they are not all presented. For the estimation procedure, the variable history length was scanned in a certain range, whereas the other past lengths were fixed to $m_x, m_y = 2$. This means that the necessary combination of history lengths, which would have been assumably enabled us to reflect the underlying connectivity, could have been missed in that constellation of embedding dimensions. Additionally, we have to keep in mind that I applied one history length parameter for all variables which are subsumed under environmental processes for the conditioned transfer entropy. As mentioned above, it is possible to assign every single environmental process a distinct past length, especially if we assume heterogeneity regarding the Markov orders of these processes. Thereby, the history length for every environmental source could be kept minimal. Otherwise, the environmental past lengths have to be equally set so that a maximal history length may ensure elimination of non-causal information transfer. But this can result in a strong underestimation of the information transfer, inducing insignificant transfer entropy.

4.6. The Limits of Causal Inference

Note that until this point, we only considered synthetic systems for information transfer. Hence, we could assume that all relevant processes in the system were completely observable. If we assume otherwise, quantifying non-causal information transfer via incomplete conditioning is unavoidable. Then, adopting maximal embedding dimension values is not only unpractical, but will also largely increase the portion of non-causal information transfer [69]. This is a main obstacle for the application of information transfer analysis to experimental data.

Generally, model SB2.2 demonstrated the very difficulty of inferring the structure of a biological system, even if suitable sampling frequencies for a transfer entropy analysis

of connectivity could be identified. We could observe that the information transfer of directionally uncoupled species, is, though lower, also significant in some cases. As the conditioned transfer entropy included all other processes in the system, the wrong choice of history lengths is a possible source of error. However, we observed that for some sampling frequencies, the increase of history lengths neither led to an increase of information transfer, nor significantly decreased non-causal information transfer. The former would be a expectable behavior if we assume information leak. Therefore, the fact that non-causal information transfer is present does not necessarily lead to an observable information leak behavior.

It is possible to argue that some systems (e.g. SB2.2) could display non-linear dynamics for which the transfer entropy does not behave similar to the information transfer observed in the context of (linear or simply non-linear) VAR processes, like the investigations in [69] suggest. In that study of Surovtsova et al., we have to consider that the topology of the analyzed systems are triplets. In my case, the systems have more than three components, and the corresponding dynamics are due to more complex non-linear kinetics. One way to approach this problem of complexity could be to divide networks into smaller parts and analyze the different motifs. Previous work demonstrated the prevalence of certain motifs in biological and artificial networks, especially of feedback and feedforward loops, which contribute to the capability of information storage [6]. For the application of transfer entropy, such a dissection of complex networks could reveal certain properties of clustered structures regarding their capability of transferring information. On the other hand, quantifying transfer entropy in smaller subsystems generates unobserved sources of information which can in turn increase non-causal information transfer.

Additionally, we have observed for model SB2.2 that the contribution of a modifying species to a reaction, which enhances the activation reaction of another species, could be considered as a symmetrical effect of the source process (modifier) on both targets (activated and inactivated form). This kinetic symmetry in the connectivity is not reflected in a comparable information transfer regarding both couplings. Therefore, in accordance with previous work about causality and statistical prediction [123, 129], information transfer, including the conditioned form, cannot be identified with structural (or mechanistic) causality in the strict sense. Rather, we should interpret transfer entropy as a measure for "functional connectivity" quantified through the statistical predictability of a target process given another information source, related to the term *functional connectivity* used in neuroscience [131, 132]. That means that the dynamics of a system can mask or reveal the underlying structural connectivity, but what it is uncovered is the statistical dependency between (sub-)systems and their spatial deployment. Such functional approaches to interactions in a network result from the interest to understand biological systems without performing perturbations, but also from the desire for model-

free concepts. We have to keep in mind that such approaches come with serious limitations if causal inference is the objective. One such restriction was introduced in Subsection 1.1.2 with the notion of complete observability referring to the work of Chicharro et al. [75]. Regarding some model used in my thesis, even complete observability did not result in a correct causal inference. As discussed above, the difficulty of finding the optimal parameter set for causal inference, as well as the estimator bias can hamper a meaningful transfer entropy analysis of connectivity.

So far, we saw how demanding the use of conditioned transfer entropy for causal inference can be. Different approaches in the past tried to shed light on the practical difference and convergence between causality and information transfer [71, 130]. According to the work of Lizier et al., in order to correctly infer causality, transfer entropy has to be used only on causal information sources for the given target (i), the measure should be conditioned to the correct set of other causal variables (ii), and the target history length has to be kept minimal (iii) [130]. These are reported as necessary, but not sufficient prerequisites for the convergence of information transfer with an inferrer of causality. Note that the first and second criterion are not met for this thesis, as non-causal sources as well as non-causal environmental processes have been used for the estimation of transfer entropy. In contrast, Surovtsova et al. show that such an inclusion of non-causal variables is not obstructive for a causal transfer entropy analysis, as long as the history lengths are set right [69]. But as mentioned before, such conclusions were made for (linear or simply non-linear) triplet systems. We should keep in mind that for larger networks with non-linear dynamics the inclusion of non-causal processes as sources or conditioned variables may introduce spurious effects in the estimation process. Especially if one cannot be sure if the embedding dimensions converge to the underlying Markov orders, such spurious effects can be increased by the inclusion of non-causal variables.
Interestingly, the third criterion of Lizier et al. suggests that a too high target history length will cause transfer entropy to deviate from its capacity to approximate causal effects. According to that, for such high history lengths, transfer entropy will evolve to a more non-causal, predictive measure which quantifies the emergent computational properties of a system on a macroscopic level. The authors show using the framework of cellular automata that transfer entropy "quantifies the information transfer by measuring the (conditional) correlation on a causal channel" [130]. In this regard, they emphasize the predictive perspective of transfer entropy which can be used as a tool to gain insights into the emergent computational structures of a system. This phenomenon is well illustrated in [130], where a specific spatial pattern of cellular automata, referred to as *gliders*, can be identified arising from the emergent properties of the system. Interestingly, Chicharro et al. advocate a similar use of transfer entropy as an indicator of the systemic

computational properties [75].
Emergent structures, which transfer information across the system, do not necessarily align with causal structures. In the context of information transfer, non-causal information flow in a network can arise from the underlying topology and mechanisms, hence the overall dynamics, and lie beyond direct and causal interactions between processes. Such non-causal information flow can be considered as emergent or computational properties of a system, especially if they appear as persistent and functional systemic properties which establish themselves upon the given direct connections and elementary mechanisms. Regarding the literature, Lizier et al. show in their work about information storage in loop motifs and clustered structures how information-theoretic measure can be used to understand biologically relevant and recurrent structures. That work can be viewed as an example for revealing the function of certain motifs which emerge from their overall structure and dynamics, and is not reducible to single causal connections. Of course, emergent information transfer will contain causal as well as non-causal information transfer. The non-causal part of information transfer can be eliminated by the conditioning on environmental processes and the right choice of parameters, e.g embedding dimensions and sampling frequency. But as we saw, this task is difficult to achieve, and I suppose that non-causal information transfer remains a part of the transfer entropy analysis. Therefore, if the conditions to infer causality cannot be met (or we have indications of non-causal information transfer), then a redirection of the information transfer analysis can be fruitful. According to Lizier et al., the embedding dimensions can be used to move the transfer entropy between a quantity which captures the microscopic effects of causality and the macroscopic impact of emergent information flow in the system [130]. Minimal embedding dimensions are more likely to capture causality, whereas maximal values detect the more emergent parts of information transfer. At this point, we conclude that if the information transfer decreases with decreasing target history length, that does not necessarily mean that non-causal information transfer is absent. In contrast, it can also imply that transfer entropy quantifies the more emergent parts of information flow in the system. In other words, if the emergent non-causal parts of information transfer are weaker than the causal parts, than by increasing the embedding dimension, transfer entropy will increasingly quantify the emergent parts, but will decrease overall. With regard to model SB2.2, this could be one explanation why we can observe decreasing information transfer upon an increase of history length even for non-coupled processes. In such cases, transfer entropy allows to identify the significant impact of processes on others beyond causal connections, which can be considered as an analysis of functional connectivity.

Additionally, instead of using global transfer entropy, which is an expected value of the whole time series, an introduction of local transfer entropy could resolve the temporal change of information flow in a system, as applied in [130]. For my thesis, I used global transfer entropy which is a weighted sum of transition probabilities over all time points used for information transfer analysis. In contrast, local transfer entropy can be used to account for the change of information transfer in time, as the single transition probabilities for each time point are not summarized. Especially, this dynamic approach to information transfer can be beneficial for cases where the dynamics change over time, as it is the case for oscillating systems. As biological signal transduction networks are often reported to exhibit oscillations [92, 133], which are also considered being functional for information encoding [88, 90], applying local transfer entropy upon oscillatory systems can give new insights into the information flow inside of their structure and information transfer to other systems, e.g. other signaling networks or gene expression.

4.7. Best Practice for Transfer Entropy Analysis

Let us consider how an evolved information transfer analysis could be performed. As mentioned before, we are left with the unsatisfactory situation that simple heuristics about the choice of parameters in the optima are not realizable. Additionally, the impact of the sampling frequency on the information transfer has to be observed. As a first step, we set a range of sampling frequencies and embedding dimensions to scan. Depending on the highest embedding dimension and number of processes to condition, a sufficiently large number of data points N should be available and fixed for the further steps. Then, we choose a criterion to maximize the reliability of predictions (e.g., locally constant prediction error, time delayed active information storage). For each sampling frequency, the chosen criterion should be calculated for the respective history lengths. Afterwards, the surfaces of the optima for the different sampling frequencies can be compared. From this point on, different approaches are possible. As a first approach, an extensive information transfer analysis can be performed where all sampling frequencies and all possible sets of embedding dimensions are used for transfer entropy estimation, if the computational resources are available. As a second approach, the range of sampling frequencies and embedding dimensions can be confined due to the results of the criterion. This will decrease the computational effort for the transfer entropy estimation, but too much confinement can result in missing the important parameters to infer functional connectivity. Subsequently, the behavior of information transfer dependent on the history

lengths can be analyzed. Sampling frequencies which exhibit information leak behavior should be excluded, as long as other sampling frequencies do not display such behavior. Otherwise, if all sampling frequencies show such information leak behavior, the range of sampling frequencies and/or embedding dimensions should be extended, or larger data lengths have to be considered. Afterwards, if adequate sampling frequencies could be identified, sets of embedding dimensions can be determined which maximize the reliability of predictions. These embedding dimensions with their respective temporal resolutions can be used for further analysis of connectivity. An evolved comparison between the results of the criterion and information transfer can be fruitful. As an additional test, a scan of different data lengths with the determined parameters can be performed to ensure that the chosen data length is in the domain of convergence or reflects the qualitative differences between variant source-target constellations sufficiently. Furthermore, the behavior of the surrogate estimates upon varying sampling frequencies should be observed. The surrogates can be used as indicators for the estimator bias, as negative transfer entropy values or surrogate information transfer above the original estimates can hint to serious numerical problems in the estimation process. Moreover, the surrogate estimates should not significantly increase with increasing embedding dimensions as this behavior suggests that temporal correlations in the surrogate data are not truly disrupted, rendering the surrogate testing invalid.

4.8. Outlook

In systems biology, the non-reductionist approach to biological networks is crucial to understand the dynamics of a system which emerge from their underlying structure and mechanisms of interaction. Especially in the context of signal transduction, tools and methods applied by systems biology can help to understand the functional properties of these information processing, encoding and transferring networks. From this perspective, the information-theoretic measure of transfer entropy is a tool to quantify the directed information transfer from one (sub-)system to another. A transfer entropy analysis can be performed upon systems to gain insights into their functional connectivity, revealing the dependencies between processes. Such dependencies can be due to causal connection, but also result from non-causal interactions in the system. For a transfer entropy analysis of connectivity in the strict sense, complete observability of the system must be given. In a highly interdependent network, the quantification of connectivity strength between processes is invalid. Rather, transfer entropy should be employed qualitatively in order

to identify the dependencies between processes. For this objective, conditioned transfer entropy can be useful to reduce or control non-causal information transfer.

Performing a transfer entropy analysis requires to assess the data and estimation dependent parameters. The data length must be sufficiently high to guarantee that qualitative differences between variant source-target constellations are observable. More specifically, source-target constellations converging to zero information transfer should be clearly discernible from constellations which are exhibiting non-zero information transfer. Importantly, the estimator bias should be decreased as much as possible. The application of an improved k-nearest neighbor estimator can attenuate the bias for estimating dependent variables [66]. Particularly for highly interdependent systems as signal transduction networks, such estimator bias hampers the information transfer analysis massively. In this context, surrogate estimations can be used as indicators of the estimator bias and contribute to identify appropriate parameters for a transfer entropy analysis of connectivity. Additionally, it is important to note that the sampling frequency of the data impacts the transfer entropy estimations significantly. Scanning different temporal resolutions for transfer entropy estimation is crucial to identify sampling frequencies for a meaningful information transfer analysis. For this, temporal resolutions exhibiting information leak behavior should be excluded from further analysis. Combining that criterion with other criteria in the literature, which maximize the reliability of predictions, is recommended for future projects applying information transfer analysis. Moreover, the sampling frequencies used in information transfer analysis can indicate the specific causal and non-causal time scales in which the processes in question interact with each other. Especially for signal transduction networks, transfer entropy analysis can be applied to decode the latency and specific time rate in which information transfer inside the networks and to other systems is realized. This can help to investigate temporal relationships between different biological networks where the effects of (de-)synchronization are important, e.g. as in neural networks or circadian rhythms. But also in order to understand the temporal correlations which generate specific dynamics in signaling pathways (oscillations, bistability), transfer entropy analysis can be promising.

As causal inference is difficult to achieve via information transfer analysis, redirecting transfer entropy to quantify the emergent properties of a system can be a fruitful approach. Especially recurrent biological motifs as feedforward and feedback loops may reveal some functional properties for signal transduction with respect to their capacity of transferring information and generating specific dependencies between processes. This can help to clarify the regulatory mechanisms in biological networks which result from such recurrent motifs and clustered structures.

Finally, local transfer entropy may be a more appropriate measure for highly dynamic systems as quantifying the change of transfer entropy in time reflects such dynamics more accurately. Most notably, the dynamics of information transfer in oscillating systems can be represented sufficiently vial local transfer entropy. That allows to understand how transfer entropy changes at different time points of the period, and if such changes of information transfer correspond to biological functions.

Bibliography

[1] Claude Elwood Shannon. "A mathematical theory of communication." In: *ACM SIGMOBILE Mobile Computing and Communications Review* 5.1 (1948), pp. 3–55.

[2] Iman Tavassoly, Joseph Goldfarb, and Ravi Iyengar. "Systems biology primer: the basic methods and approaches." In: *Essays In Biochemistry* (Oct. 4, 2018), EBC20180003. DOI: 10.1042/EBC20180003.

[3] Zaynab Mousavian, José Díaz, and Ali Masoudi-Nejad. "Information theory in systems biology. Part II: protein–protein interaction and signaling networks." In: *Seminars in Cell & Developmental Biology*. Information Theory in Systems Biology 51 (Mar. 1, 2016), pp. 14–23. DOI: 10.1016/j.semcdb.2015.12.006.

[4] Gašper Tkačik and Aleksandra M. Walczak. "Information transmission in genetic regulatory networks: a review." In: *Journal of Physics: Condensed Matter* 23.15 (Apr. 2011), p. 153102. DOI: 10.1088/0953-8984/23/15/153102.

[5] Raymond Cheong et al. "Information Transduction Capacity of Noisy Biochemical Signaling Networks." In: *Science* 334.6054 (Oct. 21, 2011), pp. 354–358. DOI: 10.1126/science.1204553.

[6] J. T. Lizier, F. M. Atay, and J. Jost. "Information storage, loop motifs, and clustered structure in complex networks." In: *Physical Review E* 86.2 (2012), p. 026110. DOI: 10.1103/physreve.86.026110.

[7] Pankaj Mehta et al. "Information processing and signal integration in bacterial quorum sensing." In: *Molecular Systems Biology* 5 (Nov. 17, 2009), p. 325. DOI: 10.1038/msb.2009.79.

[8] Etay Ziv, Ilya Nemenman, and Chris H. Wiggins. "Optimal Signal Processing in Small Stochastic Biochemical Networks." In: *PLoS ONE* 2.10 (Oct. 24, 2007). DOI: 10.1371/journal.pone.0001077.

[9] Lee Bardwell et al. "Mathematical Models of Specificity in Cell Signaling." In: *Biophysical Journal* 92.10 (May 15, 2007), pp. 3425–3441. DOI: 10.1529/biophysj.106.090084.

[10] Christian Waltermann and Edda Klipp. “Signal integration in budding yeast.” In: *Biochemical Society Transactions* 38.5 (Oct. 1, 2010), pp. 1257–1264. DOI: 10.1042/BST0381257.

[11] Narsis A. Kiani et al. “Evaluating network inference methods in terms of their ability to preserve the topology and complexity of genetic networks.” In: *Seminars in Cell & Developmental Biology*. Information Theory in Systems Biology 51 (Mar. 1, 2016), pp. 44–52. DOI: 10.1016/j.semcdb.2016.01.012.

[12] M. Wibral et al. “Revisiting Wiener’s principle of causality — interaction-delay reconstruction using transfer entropy and multivariate analysis on delay-weighted graphs.” In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Aug. 2012, pp. 3676–3679. DOI: 10.1109/EMBC.2012.6346764.

[13] Adam A Margolin et al. “ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context.” In: *BMC Bioinformatics* 7 (Suppl 1 Mar. 20, 2006), S7. DOI: 10.1186/1471-2105-7-S1-S7.

[14] Raul Vicente et al. “Transfer entropy—a model-free measure of effective connectivity for the neurosciences.” In: *Journal of Computational Neuroscience* 30.1 (Feb. 2011), pp. 45–67. DOI: 10.1007/s10827-010-0262-3.

[15] Elizabeth A. Coker et al. “SiGNet: A signaling network data simulator to enable signaling network inference.” In: *PloS One* 12.5 (2017), e0177701. DOI: 10.1371/journal.pone.0177701.

[16] Karen Sachs et al. “Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data.” In: *Science* 308.5721 (Apr. 22, 2005), pp. 523–529. DOI: 10.1126/science.1105809.

[17] Thomas Schreiber. “Measuring Information Transfer.” In: *Physical Review Letters* 85.2 (July 10, 2000), pp. 461–464. DOI: 10.1103/PhysRevLett.85.461. arXiv: nlin/0001042.

[18] Jürgen Pahle et al. “Information transfer in signaling pathways: A study using coupled simulated and experimental data.” In: *BMC Bioinformatics* 9 (Mar. 4, 2008), p. 139. DOI: 10.1186/1471-2105-9-139.

[19] Christian Widmann et al. “Mitogen-Activated Protein Kinase: Conservation of a Three-Kinase Module From Yeast to Human.” In: *Physiological Reviews* 79.1 (Jan. 1, 1999), pp. 143–180. DOI: 10.1152/physrev.1999.79.1.143.

[20] Marie Cargnello and Philippe P. Roux. "Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases." In: *Microbiology and Molecular Biology Reviews : MMBR* 75.1 (Mar. 2011), pp. 50–83. DOI: 10.1128/MMBR.00031-10.

[21] Esfandiar Maasoumi 1. "A compendium to information theory in economics and econometrics." In: *Econometric Reviews* 12.2 (Jan. 1, 1993), pp. 137–181. DOI: 10.1080/07474939308800260.

[22] Amos Golan and Esfandiar Maasoumi. "Information Theoretic and Entropy Methods: An Overview." In: *Econometric Reviews* 27.4 (May 15, 2008), pp. 317–328. DOI: 10.1080/07474930801959685.

[23] E. T. Jaynes. "Information Theory and Statistical Mechanics." In: *Physical Review* 106.4 (May 15, 1957), pp. 620–630. DOI: 10.1103/PhysRev.106.620.

[24] Y. G. Ma. "Application of Information Theory in Nuclear Liquid Gas Phase Transition." In: *Physical Review Letters* 83.18 (Nov. 1, 1999), pp. 3617–3620. DOI: 10.1103/PhysRevLett.83.3617.

[25] Philippe Réfrégier. *Noise Theory and Application to Physics: From Fluctuations to Information*. Google-Books-ID: 6oHbBwAAQBAJ. Springer Science & Business Media, Dec. 6, 2012. 294 pp.

[26] M. Bauer et al. "Finding the Direction of Disturbance Propagation in a Chemical Process Using Transfer Entropy." In: *IEEE Transactions on Control Systems Technology* 15.1 (Jan. 2007), pp. 12–21. DOI: 10.1109/TCST.2006.883234.

[27] K. R. Pattipati and M. G. Alexandridis. "Application of heuristic search and information theory to sequential fault diagnosis." In: *IEEE Transactions on Systems, Man, and Cybernetics* 20.4 (July 1990), pp. 872–887. DOI: 10.1109/21.105086.

[28] Robert F. Wagner, David G. Brown, and Mary S. Pastel. "Application of information theory to the assessment of computed tomography." In: *Medical Physics* 6.2 (Mar. 1, 1979), pp. 83–94. DOI: 10.1118/1.594559.

[29] N. A. Brunsell. "A multiscale information theory approach to assess spatial–temporal variability of daily precipitation." In: *Journal of Hydrology* 385.1 (May 7, 2010), pp. 165–172. DOI: 10.1016/j.jhydrol.2010.02.016.

[30] B. Ya. Ryabko and V. A. Monarev. "Using information theory approach to randomness testing." In: *Journal of Statistical Planning and Inference* 133.1 (July 1, 2005), pp. 95–110. DOI: 10.1016/j.jspi.2004.02.010.

[31] R. A. Johnson. "An Information Theory Approach to Diagnosis." In: *IRE Transactions on Reliability and Quality Control* RQC-9.1 (Apr. 1960), pp. 35–35. DOI: 10.1109/IRE-PGRQC.1960.5007263.

[32] Harold Wolman and Ed Page. "Policy Transfer among Local Governments: An Information–Theory Approach." In: *Governance* 15.4 (2002), pp. 577–501. DOI: 10.1111/1468-0491.00198.

[33] Hubert P. Yockey. "An application of information theory to the central dogma and the sequence hypothesis." In: *Journal of Theoretical Biology* 46.2 (Aug. 1, 1974), pp. 369–406. DOI: 10.1016/0022-5193(74)90005-8.

[34] S. P. Strong et al. "On the application of information theory to neural spike trains." In: *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (1998), pp. 621–632.

[35] Ilya Nemenman. "Information theory, multivariate dependence, and genetic network inference." In: *arXiv:q-bio/0406015* (June 7, 2004). arXiv: q-bio/0406015.

[36] Hasan Metin Aktulga et al. "Identifying Statistical Dependence in Genomic Sequences via Mutual Information Estimates." In: *EURASIP J. Bioinformatics Syst. Biol.* 2007 (July 2007), 3:1–3:11. DOI: 10.1155/2007/14741.

[37] Alex Rhee, Raymond Cheong, and Andre Levchenko. "The application of information theory to biochemical signaling systems." In: *Physical Biology* 9.4 (Aug. 2012), p. 045011. DOI: 10.1088/1478-3975/9/4/045011.

[38] Hector Zenil, Narsis A. Kiani, and Jesper Tegnér. "Methods of information theory and algorithmic complexity for network biology." In: *Seminars in Cell & Developmental Biology*. Information Theory in Systems Biology 51 (Mar. 1, 2016), pp. 32–43. DOI: 10.1016/j.semcdb.2016.01.011.

[39] Pietro Zoppoli, Sandro Morganella, and Michele Ceccarelli. "TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach." In: *BMC Bioinformatics* 11.1 (Mar. 25, 2010), p. 154. DOI: 10.1186/1471-2105-11-154.

[40] Miguel Lopes and Gianluca Bontempi. "Experimental assessment of static and dynamic algorithms for gene regulation inference from time series expression data." In: *Frontiers in Genetics* 4 (2013). DOI: 10.3389/fgene.2013.00303.

[41] Alexander G. Dimitrov, Aurel A. Lazar, and Jonathan D. Victor. "Information theory in neuroscience." In: *Journal of Computational Neuroscience* 30.1 (Feb. 1, 2011), pp. 1–5. DOI: 10.1007/s10827-011-0314-3.

[42] O. Milenkovic et al. “Introduction to the Special Issue on Information Theory in Molecular Biology and Neuroscience.” In: *IEEE Transactions on Information Theory* 56.2 (Feb. 2010), pp. 649–652. DOI: 10.1109/TIT.2009.2036971.

[43] Joseph T. Lizier et al. “Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity.” In: *Journal of Computational Neuroscience* 30.1 (Feb. 1, 2011), pp. 85–107. DOI: 10.1007/s10827-010-0271-2.

[44] C. E. Shannon and J. McCarthy. *Automata Studies. (AM-34)*. Google-Books-ID: adLfCwAAQBAJ. Princeton University Press, 1956. 297 pp.

[45] S Winograd, Herman H Goldstine, and J. D Cowan. *Reliable computation in the presence of noise*. OCLC: 1668759. Cambridge, Mass.: M.I.T. Press, 1963.

[46] T. M Cover and Joy A Thomas. *Elements of information theory*. OCLC: 22305429. New York: Wiley, 1991.

[47] Thomas D. Schneider. “Theory of molecular machines. I. Channel capacity of molecular machines.” In: *Journal of Theoretical Biology* 148.1 (Jan. 7, 1991), pp. 83–123. DOI: 10.1016/S0022-5193(05)80466-7.

[48] Terry Bossomaier et al. *An Introduction to Transfer Entropy*. Cham: Springer International Publishing, 2016. DOI: 10.1007/978-3-319-43222-9.

[49] Michael Wibral, Raul Vicente, and Michael Lindner. “Transfer Entropy in Neuroscience.” In: *Directed Information Measures in Neuroscience*. Ed. by Michael Wibral, Raul Vicente, and Joseph T. Lizier. Understanding Complex Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 3–36. DOI: 10.1007/978-3-642-54474-3_1.

[50] Michael Wibral et al. “Measuring Information-Transfer Delays.” In: *PLOS ONE* 8.2 (Feb. 28, 2013), e55809. DOI: 10.1371/journal.pone.0055809.

[51] Patricia Wollstadt et al. “Efficient Transfer Entropy Analysis of Non-Stationary Neural Time Series.” In: *PLOS ONE* 9.7 (July 28, 2014), e102833. DOI: 10.1371/journal.pone.0102833.

[52] C. W. J. Granger. “Investigating Causal Relations by Econometric Models and Cross-spectral Methods.” In: *Econometrica* 37.3 (1969), pp. 424–438. DOI: 10.2307/1912791.

[53] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. “Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables.” In: *Physical Review Letters* 103.23 (Dec. 4, 2009), p. 238701. DOI: 10.1103/PhysRevLett.103.238701.

[54] Katerina Hlaváčková-Schindler. “Equivalence of Granger Causality and Transfer Entropy: A Generalization.” In: 5 (Jan. 2011), pp. 3637–3648.

[55] Immo Weber et al. “The influence of filtering and downsampling on the estimation of transfer entropy.” In: *PLOS ONE* 12.11 (Nov. 17, 2017), e0188210. DOI: 10.1371/journal.pone.0188210.

[56] Sourabh Lahiri et al. “Information-theoretic analysis of the directional influence between cellular processes.” In: *PLOS ONE* 12.11 (Nov. 9, 2017), e0187431. DOI: 10.1371/journal.pone.0187431.

[57] Arne Schoch. “Stochastic Simulation and Information-theoretic Analysis of the Calcium-dependent Activation of Protein Kinase C.” Master Thesis. Heidelberg: University of Heidelberg, 2016.

[58] A. Kaiser and T. Schreiber. “Information transfer in continuous processes.” In: *Physica D: Nonlinear Phenomena* 166.1 (June 1, 2002), pp. 43–62. DOI: 10.1016/S0167-2789(02)00432-3.

[59] Peter Grassberger. “Finite sample corrections to entropy and dimension estimates.” In: *Physics Letters A* 128.6 (Apr. 11, 1988), pp. 369–373. DOI: 10.1016/0375-9601(88)90193-4.

[60] Murray Rosenblatt. “Remarks on Some Nonparametric Estimates of a Density Function.” In: *The Annals of Mathematical Statistics* 27.3 (Sept. 1956), pp. 832–837. DOI: 10.1214/aoms/1177728190.

[61] Emanuel Parzen. “On Estimation of a Probability Density Function and Mode.” In: *The Annals of Mathematical Statistics* 33.3 (Sept. 1962), pp. 1065–1076. DOI: 10.1214/aoms/1177704472.

[62] J. M. Nichols et al. “Detecting nonlinearity in structural systems using the transfer entropy.” In: *Physical Review E* 72.4 (Oct. 24, 2005), p. 046217. DOI: 10.1103/PhysRevE.72.046217.

[63] Joon Lee et al. “Transfer Entropy Estimation and Directional Coupling Change Detection in Biomedical Time Series.” In: *BioMedical Engineering OnLine* 11.1 (Apr. 13, 2012), p. 19. DOI: 10.1186/1475-925X-11-19.

[64] Alexander Kraskov, Harald Stoegbauer, and Peter Grassberger. “Estimating Mutual Information.” In: *Physical Review E* 69.6 (June 23, 2004). DOI: 10.1103/PhysRevE.69.066138. arXiv: cond-mat/0305641.

[65] Shiraj Khan et al. “Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data.” In: *Physical Review E* 76.2 (Aug. 14, 2007), p. 026209. DOI: 10.1103/PhysRevE.76.026209.

[66] Shuyang Gao, Greg Ver Steeg, and Aram Galstyan. "Efficient Estimation of Mutual Information for Strongly Dependent Variables." In: *arXiv:1411.2003 [physics, stat]* (Nov. 7, 2014). arXiv: 1411.2003.

[67] Shashank Singh and Barnabás Póczos. "Analysis of k-Nearest Neighbor Distances with Application to Entropy Estimation." In: *arXiv:1603.08578 [cs, math, stat]* (Mar. 28, 2016). arXiv: 1603.08578.

[68] James Theiler et al. "Testing for nonlinearity in time series: the method of surrogate data." In: *Physica D: Nonlinear Phenomena* 58.1 (Sept. 15, 1992), pp. 77–94. DOI: 10.1016/0167-2789(92)90102-S.

[69] Irina Surovtsova and Jürgen Pahle. "Transfer Entropy Analysis with VAR processes." draft. Bioquant, Heidelberg, 2019.

[70] D. Chicharro and A. Ledberg. "Framework to study dynamic dependencies in networks of interacting processes." In: *Phys. Rev. E* 86.4 (2012), p. 041901. DOI: 10.1103/PhysRevE.86.041901.

[71] Dmitry A. Smirnov. "Spurious causalities with transfer entropy." In: *Physical Review E* 87.4 (Apr. 17, 2013). DOI: 10.1103/PhysRevE.87.042917.

[72] Vasily A. Vakorin, Olga A. Krakovska, and Anthony R. McIntosh. "Confounding effects of indirect connections on causality estimation." In: *Journal of Neuroscience Methods* 184.1 (Oct. 30, 2009), pp. 152–160. DOI: 10.1016/j.jneumeth.2009.07.014.

[73] David Ruelle and Floris Takens. "On the nature of turbulence." In: *Communications in Mathematical Physics* 20.3 (Sept. 1, 1971), pp. 167–192. DOI: 10.1007/BF01646553.

[74] Mario Ragwitz and Holger Kantz. "Markov models from data by simple nonlinear time series predictors in delay embedding spaces." In: *Physical Review E* 65.5 (Apr. 15, 2002), p. 056201. DOI: 10.1103/PhysRevE.65.056201.

[75] Daniel Chicharro and Anders Ledberg. "When Two Become One: The Limits of Causality Analysis of Brain Dynamics." In: *PLoS ONE* 7.3 (Mar. 16, 2012). Ed. by Thomas Wennekers, e32466. DOI: 10.1371/journal.pone.0032466.

[76] Shinsuke Uda and Shinya Kuroda. "Analysis of cellular signal transduction from an information theoretic approach." In: *Seminars in Cell & Developmental Biology*. Information Theory in Systems Biology 51 (Mar. 1, 2016), pp. 24–31. DOI: 10.1016/j.semcdb.2015.12.011.

[77] Christian Waltermann and Edda Klipp. "Information theory based approaches to cellular signaling." In: *Biochimica et Biophysica Acta (BBA) - General Subjects*. Systems Biology of Microorganisms 1810.10 (Oct. 1, 2011), pp. 924–932. DOI: 10.1016/j.bbagen.2011.07.009.

[78] Sahand Hormoz. "Cross Talk and Interference Enhance Information Capacity of a Signaling Pathway." In: *Biophysical Journal* 104.5 (Mar. 5, 2013), pp. 1170–1180. DOI: 10.1016/j.bpj.2013.01.033.

[79] Toby Berger. "Rate-Distortion Theory." In: *Wiley Encyclopedia of Telecommunications*. American Cancer Society, 2003. DOI: 10.1002/0471219282.eot142.

[80] Joshua R. Porter, Burton W. Andrews, and Pablo A. Iglesias. "A framework for designing and analyzing binary decision-making strategies in cellular systems." In: *Integrative biology : quantitative biosciences from nano to macro* 4.3 (Mar. 2012), pp. 310–317. DOI: 10.1039/C2IB90009B.

[81] Burton W Andrews and Pablo A Iglesias. "An Information-Theoretic Characterization of the Optimal Gradient Sensing Response of Cells." In: *PLoS Computational Biology* 3.8 (Aug. 2007). DOI: 10.1371/journal.pcbi.0030153.

[82] R. J. Davis. "The mitogen-activated protein kinase signal transduction pathway." In: *Journal of Biological Chemistry* 268.20 (July 15, 1993), pp. 14553–14556.

[83] C Y Huang and J E Ferrell. "Ultrasensitivity in the mitogen-activated protein kinase cascade." In: *Proceedings of the National Academy of Sciences of the United States of America* 93.19 (Sept. 17, 1996), pp. 10078–10083.

[84] Sung-Young Shin et al. "Positive- and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-MEK-ERK signal transduction pathway." In: *Journal of Cell Science* 122.3 (Feb. 1, 2009), pp. 425–435. DOI: 10.1242/jcs.036319.

[85] James E Ferrell. "Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability." In: *Current Opinion in Cell Biology* 14.2 (Apr. 1, 2002), pp. 140–148. DOI: 10.1016/S0955-0674(02)00314-9.

[86] Boris N. Kholodenko. "Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades: Sustained oscillations in MAPK cascades." In: *European Journal of Biochemistry* 267.6 (Mar. 2000), pp. 1583–1588. DOI: 10.1046/j.1432-1327.2000.01197.x.

[87] Uddipan Sarma and Indira Ghosh. "Oscillations in MAPK cascade triggered by two distinct designs of coupled positive and negative feedback loops." In: *BMC Research Notes* 5 (June 13, 2012), p. 287. DOI: 10.1186/1756-0500-5-287.

[88] Harish Shankaran and H Steven Wiley. "Oscillatory dynamics of the extracellular signal-regulated kinase pathway." In: *Current Opinion in Genetics & Development*. Genetics of system biology 20.6 (Dec. 1, 2010), pp. 650–655. DOI: 10.1016/j.gde.2010.08.002.

[89] Vijay Chickarmane, Boris N. Kholodenko, and Herbert M. Sauro. "Oscillatory dynamics arising from competitive inhibition and multisite phosphorylation." In: *Journal of Theoretical Biology* 244.1 (Jan. 7, 2007), pp. 68–76. DOI: 10.1016/j.jtbi.2006.05.013.

[90] Raymond Cheong and Andre Levchenko. "Oscillatory signaling processes: the how, the why and the where." In: *Current opinion in genetics & development* 20.6 (Dec. 2010), pp. 665–669. DOI: 10.1016/j.gde.2010.08.007.

[91] Dirk Fey et al. "The complexities and versatility of the RAS-to-ERK signalling system in normal and cancer cells." In: *Seminars in Cell & Developmental Biology*. Cardiac Regeneration 58 (Oct. 1, 2016), pp. 96–107. DOI: 10.1016/j.semcdb.2016.06.011.

[92] Yaman Arkun and Mohammadreza Yasemi. "Dynamics and control of the ERK signaling pathway: Sensitivity, bistability, and oscillations." In: *PLoS ONE* 13.4 (Apr. 9, 2018). DOI: 10.1371/journal.pone.0195513.

[93] Reinhart Heinrich and Tom A. Rapoport. "A Linear Steady-State Treatment of Enzymatic Chains." In: *European Journal of Biochemistry* 42.1 (1974), pp. 89–95. DOI: 10.1111/j.1432-1033.1974.tb03318.x.

[94] H. Kacser et al. "The control of flux." In: *Biochemical Society Transactions* 23.2 (May 1, 1995), pp. 341–366. DOI: 10.1042/bst0230341.

[95] Daniel T. Gillespie. "Exact stochastic simulation of coupled chemical reactions." In: *The Journal of Physical Chemistry* 81.25 (Dec. 1, 1977), pp. 2340–2361. DOI: 10.1021/j100540a008.

[96] Daniel T Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions." In: *Journal of Computational Physics* 22.4 (Dec. 1, 1976), pp. 403–434. DOI: 10.1016/0021-9991(76)90041-3.

[97] Adam Arkin, John Ross, and Harley H. McAdams. "Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage -Infected Escherichia coli Cells." In: *Genetics* 149.4 (Aug. 1, 1998), pp. 1633–1648.

[98] Ian L. Ross, Catherine M. Browne, and David A. Hume. "Transcription of individual genes in eukaryotic cells occurs randomly and infrequently." In: *Immunology & Cell Biology* 72.2 (1994), pp. 177–185. DOI: 10.1038/icb.1994.26.

[99] Christopher V. Rao, Denise M. Wolf, and Adam P. Arkin. "Control, exploitation and tolerance of intracellular noise." In: *Nature* 420.6912 (Nov. 2002), p. 231. DOI: 10.1038/nature01258.

[100] David A. Hume. "Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression." In: *Blood* 96.7 (Oct. 1, 2000), pp. 2323–2328.

[101] Liqing Wang, Inanç Birol, and Vassily Hatzimanikatis. "Metabolic Control Analysis under Uncertainty: Framework Development and Case Studies." In: *Biophysical Journal* 87.6 (Dec. 1, 2004), pp. 3750–3763. DOI: 10.1529/biophysj.104.048090.

[102] Rudiyanto Gunawan et al. "Sensitivity Analysis of Discrete Stochastic Systems." In: *Biophysical Journal* 88.4 (Apr. 2005), pp. 2530–2540. DOI: 10.1529/biophysj.104.053405.

[103] Eric L. Haseltine and James B. Rawlings. "Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics." In: *The Journal of Chemical Physics* 117.15 (Sept. 30, 2002), pp. 6959–6969. DOI: 10.1063/1.1505860.

[104] Jürgen Pahle. "Stochastic simulation and analysis of biochemical networks." In: (June 27, 2008). DOI: http://dx.doi.org/10.18452/15786.

[105] Frank Wilcoxon. "Individual Comparisons by Ranking Methods." In: *Biometrics Bulletin* 1.6 (1945), pp. 80–83. DOI: 10.2307/3001968.

[106] Yosef Hochberg. "A sharper Bonferroni procedure for multiple tests of significance." In: *Biometrika* 75.4 (Dec. 1, 1988), pp. 800–802. DOI: 10.1093/biomet/75.4.800.

[107] Sture Holm. "A simple sequentially rejective multiple test procedure." In: *Scandinavian Journal of Statistics. Theory and Applications* 6.2 (1979), pp. 65–70.

[108] Stefan Hoops et al. "COPASI—a COmplex PAthway SImulator." In: *Bioinformatics* 22.24 (Dec. 15, 2006), pp. 3067–3074. DOI: 10.1093/bioinformatics/btl485.

[109] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019.

[110] Jonas Förster and Jürgen Pahle. *CoRC: COPASI R Connector*. 2019.

[111] Hadley Wickham and Lionel Henry. *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. 2019.

[112] Hadley Wickham. "The Split-Apply-Combine Strategy for Data Analysis." In: *Journal of Statistical Software* 40.1 (2011), pp. 1–29.

[113] Hadley Wickham et al. *dplyr: A Grammar of Data Manipulation*. 2019.

[114] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[115] Ionut Florescu. *Probability and Stochastic Processes*. Google-Books-ID: Kdq6BQAAQBAJ. John Wiley & Sons, Dec. 4, 2014. 572 pp.

[116] Deniz Gençağa. "Transfer Entropy." In: *Entropy* 20.4 (Apr. 16, 2018), p. 288. DOI: 10.3390/e20040288.

[117] Lionel Barnett and Anil K. Seth. "Detectability of Granger causality for subsampled continuous-time neurophysiological processes." In: *Journal of Neuroscience Methods* 275 (Jan. 1, 2017), pp. 93–121. DOI: 10.1016/j.jneumeth.2016.10.016.

[118] Joseph T. Lizier. "JIDT: An Information-Theoretic Toolkit for Studying the Dynamics of Complex Systems." In: *Frontiers in Robotics and AI* 1 (2014). DOI: 10.3389/frobt.2014.00011.

[119] Raul Vicente and Michael Wibral. "Efficient Estimation of Information Transfer." In: *Directed Information Measures in Neuroscience*. Ed. by Michael Wibral, Raul Vicente, and Joseph T. Lizier. Understanding Complex Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 37–58. DOI: 10.1007/978-3-642-54474-3_2.

[120] Jörg Breitung and Norman R. Swanson. "Temporal aggregation and spurious instantaneous causality in multiple time series models." In: *Journal of Time Series Analysis* 23.6 (2002), pp. 651–665. DOI: 10.1111/1467-9892.00284.

[121] Esther Florin et al. "The effect of filtering on Granger causality based multivariate causality measures." In: *NeuroImage* 50.2 (Apr. 1, 2010), pp. 577–588. DOI: 10.1016/j.neuroimage.2009.12.050.

[122] D. A. Smirnov and B. P. Bezruchko. "Spurious causalities due to low temporal resolution: Towards detection of bidirectional coupling from time series." In: *EPL (Europhysics Letters)* 100.1 (Oct. 2012), p. 10005. DOI: 10.1209/0295-5075/100/10005.

[123] Lionel Barnett and Terry Bossomaier. "Transfer Entropy as a Log-Likelihood Ratio." In: *Physical Review Letters* 109.13 (Sept. 28, 2012), p. 138105. DOI: 10.1103/PhysRevLett.109.138105.

[124] Koichi Takahashi, Sorin Tănase-Nicola, and Pieter Rein ten Wolde. "Spatio-temporal correlations can drastically change the response of a MAPK pathway." In: *Proceedings of the National Academy of Sciences* 107.6 (Feb. 9, 2010), pp. 2473–2478. DOI: 10.1073/pnas.0906885107.

[125] Joshua Garland, Ryan G. James, and Elizabeth Bradley. "Leveraging information storage to select forecast-optimal parameters for delay-coordinate reconstructions." In: *Physical Review E* 93.2 (Feb. 29, 2016), p. 022221. DOI: 10.1103/PhysRevE.93.022221.

[126] C. W. J. Granger. "Testing for causality: A personal viewpoint." In: *Journal of Economic Dynamics and Control* 2 (Jan. 1, 1980), pp. 329–352. DOI: 10.1016/0165-1889(80)90069-X.

[127] Michael Eichler. "Causal Inference in Time Series Analysis." In: *Wiley Series in Probability and Statistics*. Ed. by Carlo Berzuini, Philip Dawid, and Luisa Bernardinelli. Chichester, UK: John Wiley & Sons, Ltd, June 25, 2012, pp. 327–354. DOI: 10.1002/9781119945710.ch22.

[128] Anna Krakovská et al. "Comparison of six methods for the detection of causality in a bivariate time series." In: *Physical Review E* 97.4 (Apr. 10, 2018), p. 042207. DOI: 10.1103/PhysRevE.97.042207.

[129] Steven L. Bressler and Anil K. Seth. "Wiener–Granger Causality: A well established methodology." In: *NeuroImage* 58.2 (Sept. 15, 2011), pp. 323–329. DOI: 10.1016/j.neuroimage.2010.02.059.

[130] J. T. Lizier and M. Prokopenko. "Differentiating information transfer and causal effect." In: *The European Physical Journal B* 73.4 (Feb. 2010), pp. 605–615. DOI: 10.1140/epjb/e2010-00034-5.

[131] Rainer Goebel et al. "Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping." In: *Magnetic Resonance Imaging* 21.10 (Dec. 1, 2003), pp. 1251–1261. DOI: 10.1016/j.mri.2003.08.026.

[132] Alard Roebroeck, Elia Formisano, and Rainer Goebel. "The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution." In: *NeuroImage* 58.2 (Sept. 15, 2011), pp. 296–302. DOI: 10.1016/j.neuroimage.2009.09.036.

[133] U Kummer et al. "Switching from simple to complex oscillations in calcium signaling." In: *Biophysical Journal* 79.3 (Sept. 2000), pp. 1188–1195.

Appendices

A. Additional Results

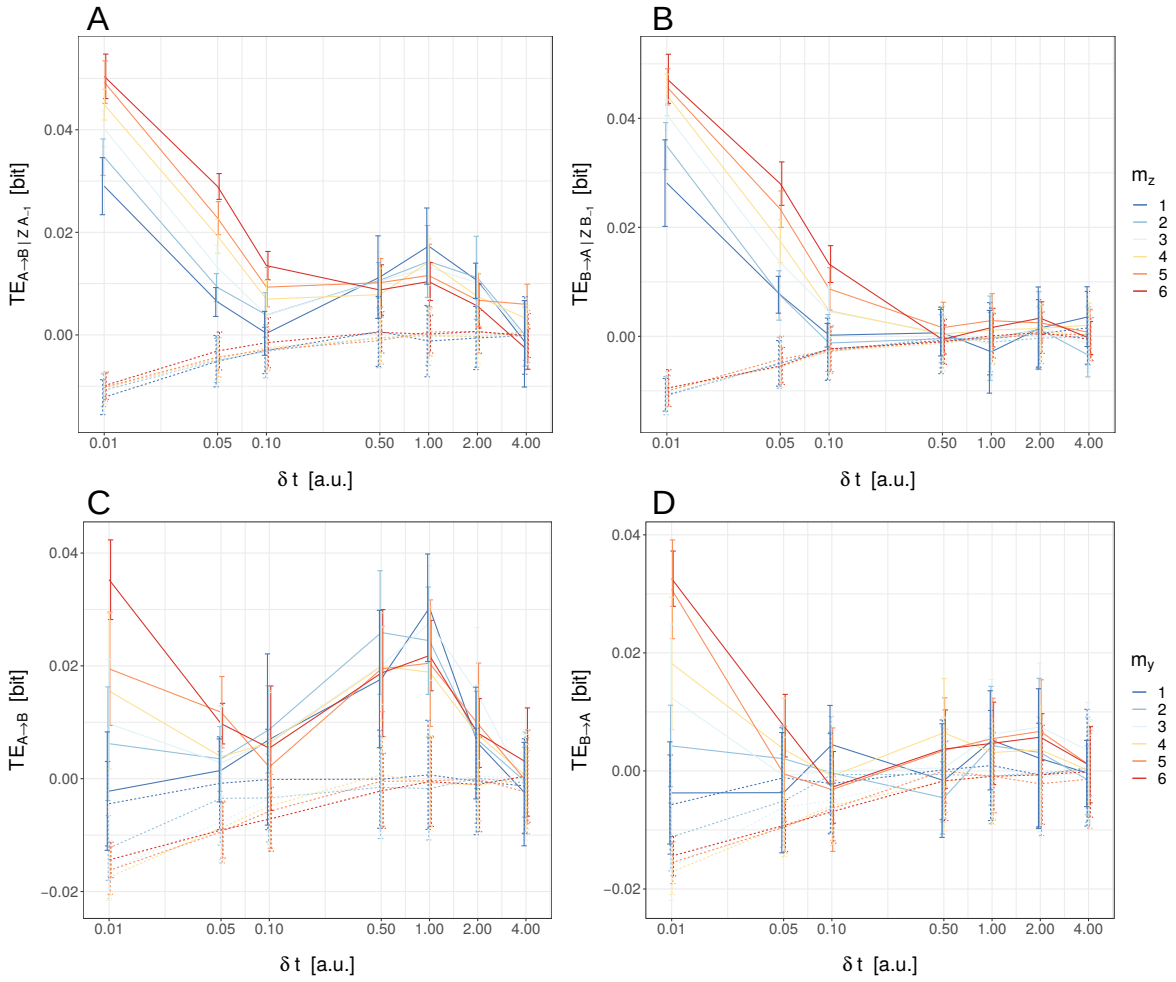

Figure A.1: δt-scan for model SL1.2. TE estimations in dependency of the sampling frequency δt with varying environmental and source history lengths m_z and m_y respectively are shown. Information transfer for different source-target constellations between the species of model SL1.2 are computed: **A,C:** $A \to B$, **B,D:** $B \to A$. Among these, only $A \to B$ is a truly connected coupling. The upper part displays conditioned, the lower part unconditioned TE. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original or 100 surrogate estimates respectively). Computations are realized with data length $N = 5000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.

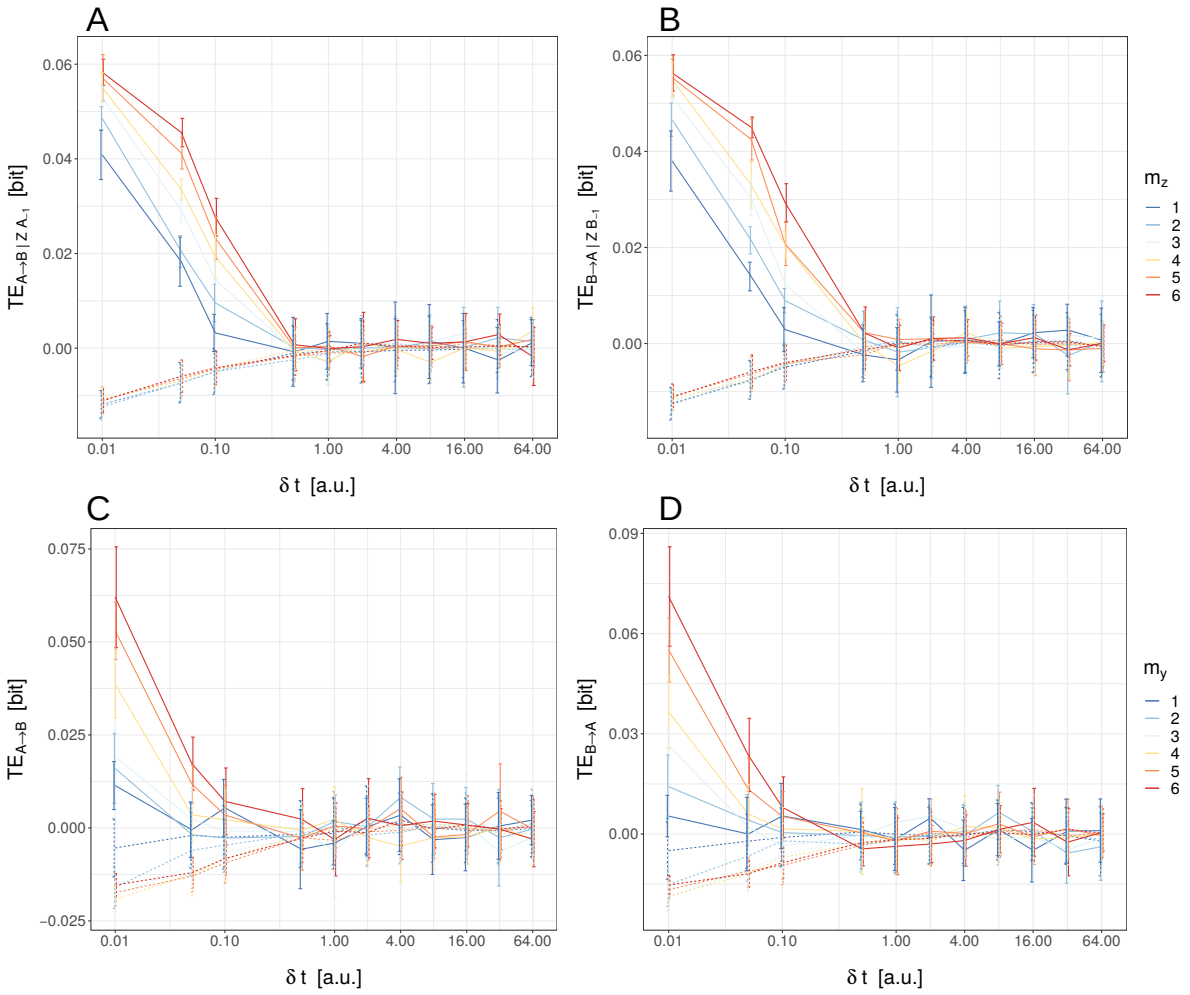

Figure A.2: δt-scan for model SB1.2. TE estimations in dependency of the sampling frequency δt with varying environmental and source history lengths m_z and m_y respectively are shown. Information transfer for different source-target constellations between the species of model SB1.2 are computed: **A,C:** $A \to B$, **B,D:** $B \to A$. Among these, only $A \to B$ is a truly connected coupling. The upper part displays conditioned, the lower part unconditioned TE. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original or 100 surrogate estimates respectively). Computations are realized with data length $N = 5000$ and $m_x, m_y = 2$ for the unvarying embedding dimensions. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.

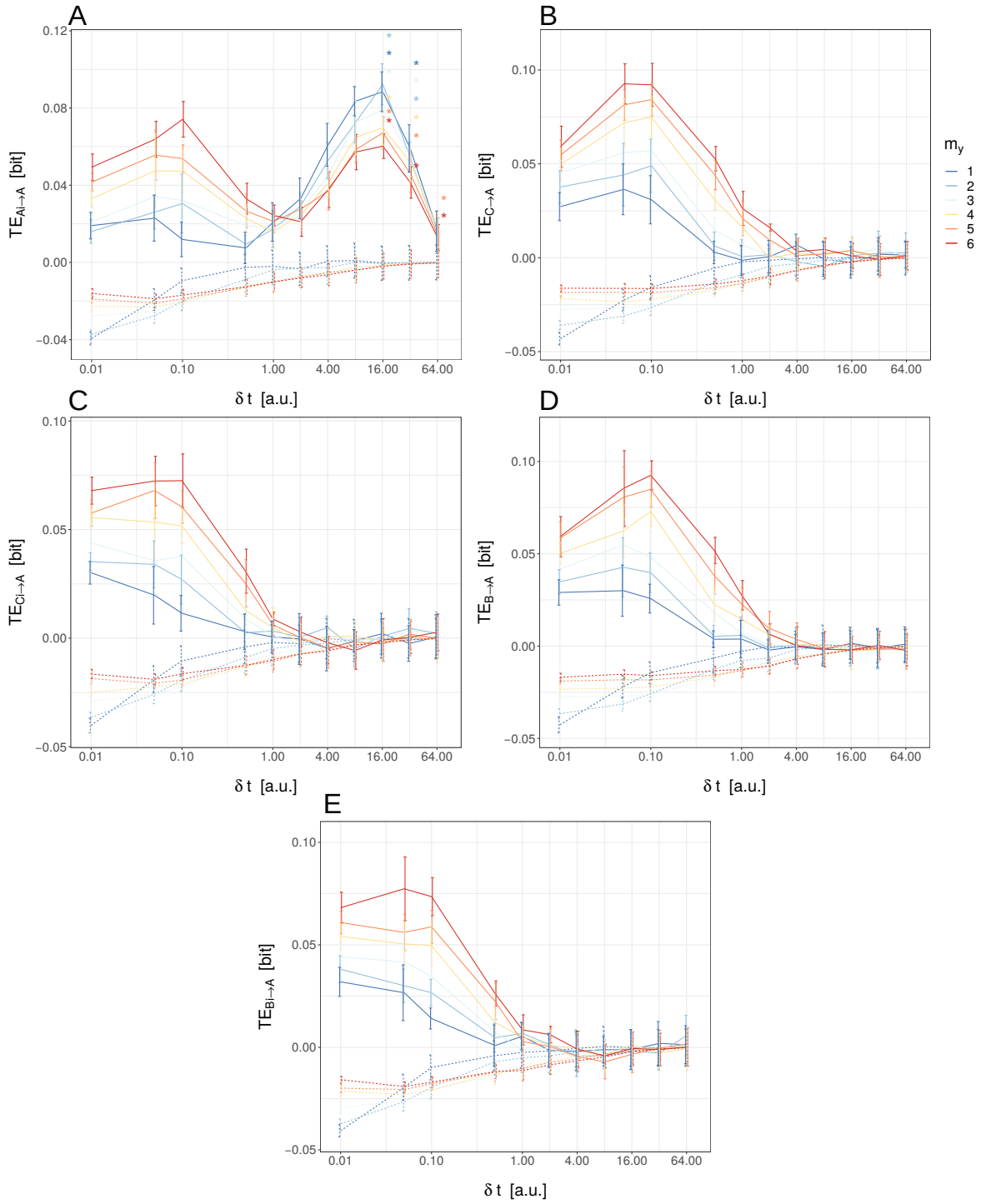

Figure A.3: δt-scan for model SB2. Unconditioned TE estimations in dependency of the sampling frequency δt and source history length m_y for the target species A are shown. Information transfer for different source-target constellations between the species of model SB2 are computed: **A:** $A_i \to A$, **B:** $C \to A$, **C:** $C_i \to A$, **D:** $B \to A$, **E:** $B_i \to A$. Among these, only panel **(A)** and **(B)** display truly connected couplings. Dashed lines indicate the respective surrogate estimations. Asterisks show significance of original estimates with difference to surrogates (one-sample Wilcoxon signed rank test with $p \leqslant 0.01$, Holm-Bonferroni corrected). Note that the significance level is only calculated for δt 16, 32 and 64 [a.u.] with 20 original estimates per data point, where 1000 surrogate estimates correspond to one original estimate. Otherwise, for every original estimate (with a total of 10 original estimates), 10 surrogate estimation were performed, and no significance calculation was applied. Computations are realized with data length $N = 5000$ and $m_x = 2$ for the unvarying embedding dimension. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

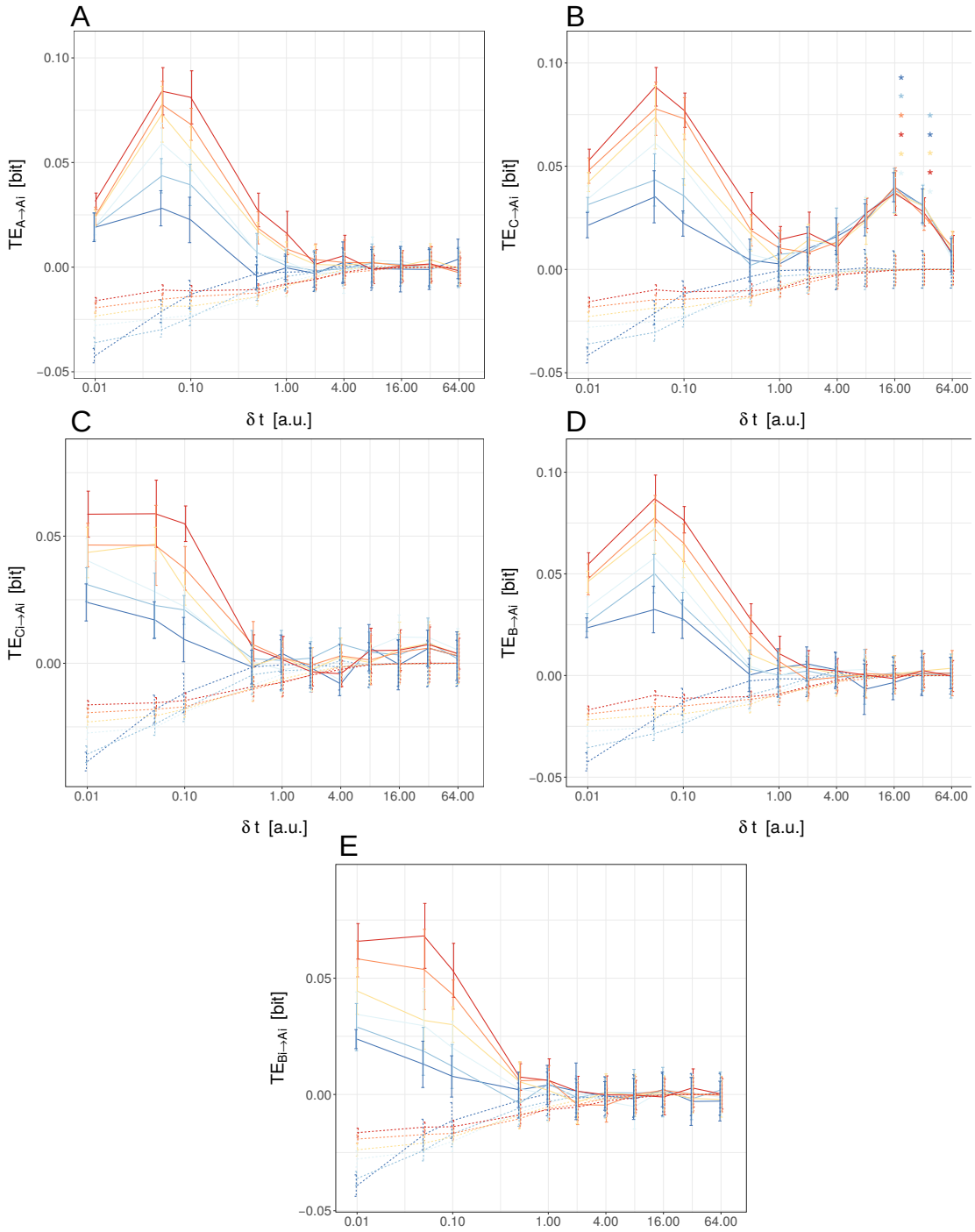

Figure A.4: δt-scan for model SB2. Unconditioned TE estimations in dependency of the sampling frequency δt and source history length m_y for the target species A_i are shown. Information transfer for different source-target constellations between the species of model SB2 are computed: **A:** $A \to A_i$, **B:** $C \to A_i$, **C:** $C_i \to A_i$, **D:** $B \to A_i$, **E:** $B_i \to A_i$. Among these, only panel **(B)** displays a truly connected coupling. Dashed lines indicate the respective surrogate estimations. Asterisks show significance of original estimates with difference to surrogates (one-sample Wilcoxon signed rank test with $p \leqslant 0.01$, Holm-Bonferroni corrected). Note that the significance level is only calculated for δt 16, 32 and 64 [a.u.] with 20 original estimates per data point, where 1000 surrogate estimates correspond to one original estimate. Otherwise, for every original estimate (with a total of 10 original estimates), 10 surrogate estimation were performed, and no significance calculation was applied. Computations are realized with data length $N = 5000$ and $m_x = 2$ for the unvarying embedding dimension. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

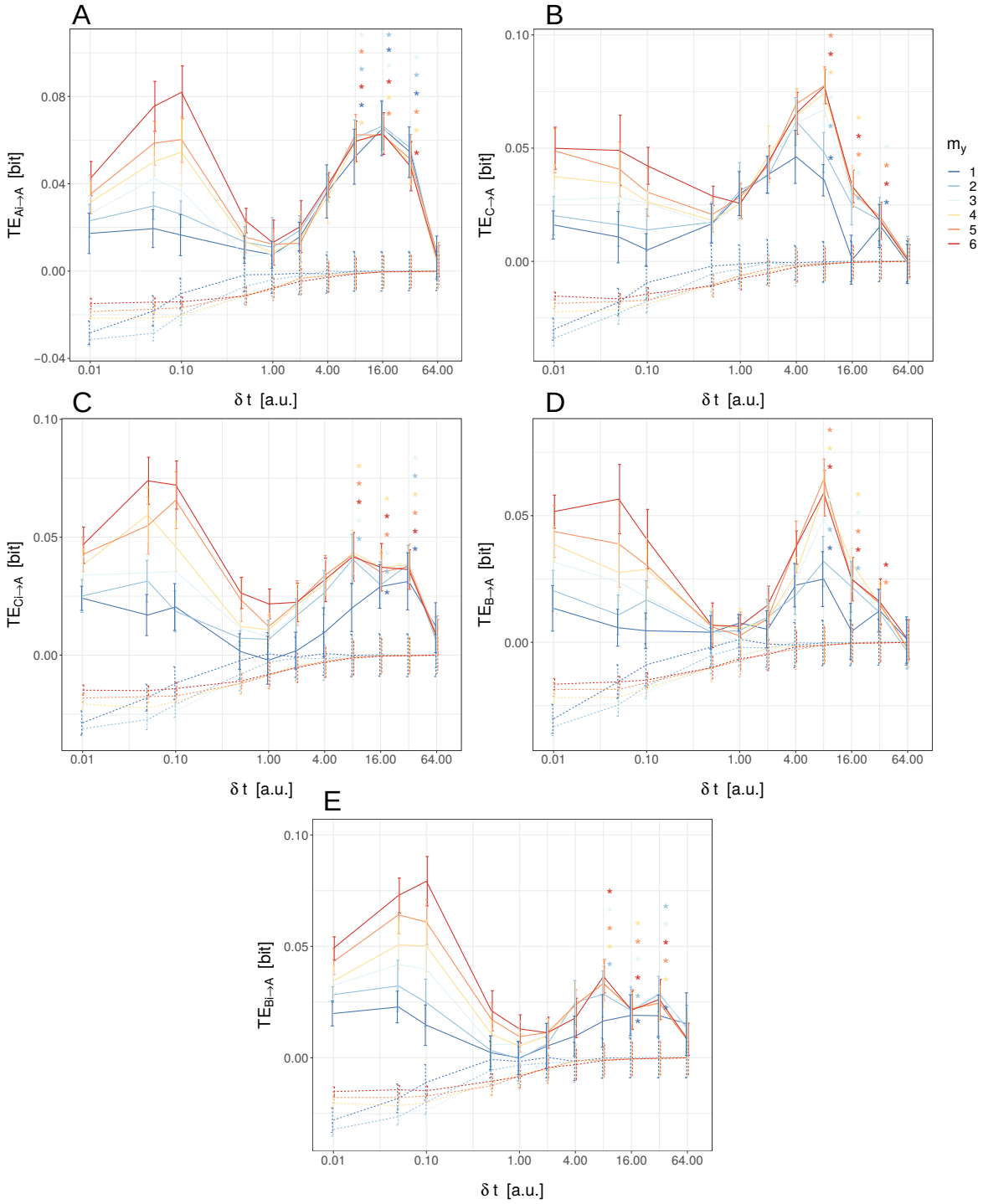

Figure A.5: δt-scan for model SB2.2. Unconditioned TE estimations in dependency of the sampling frequency δt and source history length m_y for the target species A are shown. Information transfer for different source-target constellations between the species of model SB2.2 are computed: **A:** $A_i \to A$, **B:** $C \to A$, **C:** $C_i \to A$, **D:** $B \to A$, **E:** $B_i \to A$. Among these, only panel **(A)** and **(B)** display truly connected couplings. Dashed lines indicate the respective surrogate estimations. Asterisks show significance of original estimates with difference to surrogates (one-sample Wilcoxon signed rank test with $p \leqslant 0.01$, Holm-Bonferroni corrected). Note that the significance level is only calculated for δt 8, 16, 32 and 64 [a.u.] with 20 original estimates per data point, where 1000 surrogate estimates correspond to one original estimate. Otherwise, for every original estimate (with a total of 10 original estimates), 10 surrogate estimation were performed, and no significance calculation was applied. Computations are realized with data length $N = 5000$ and $m_x = 2$ for the unvarying embedding dimension. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

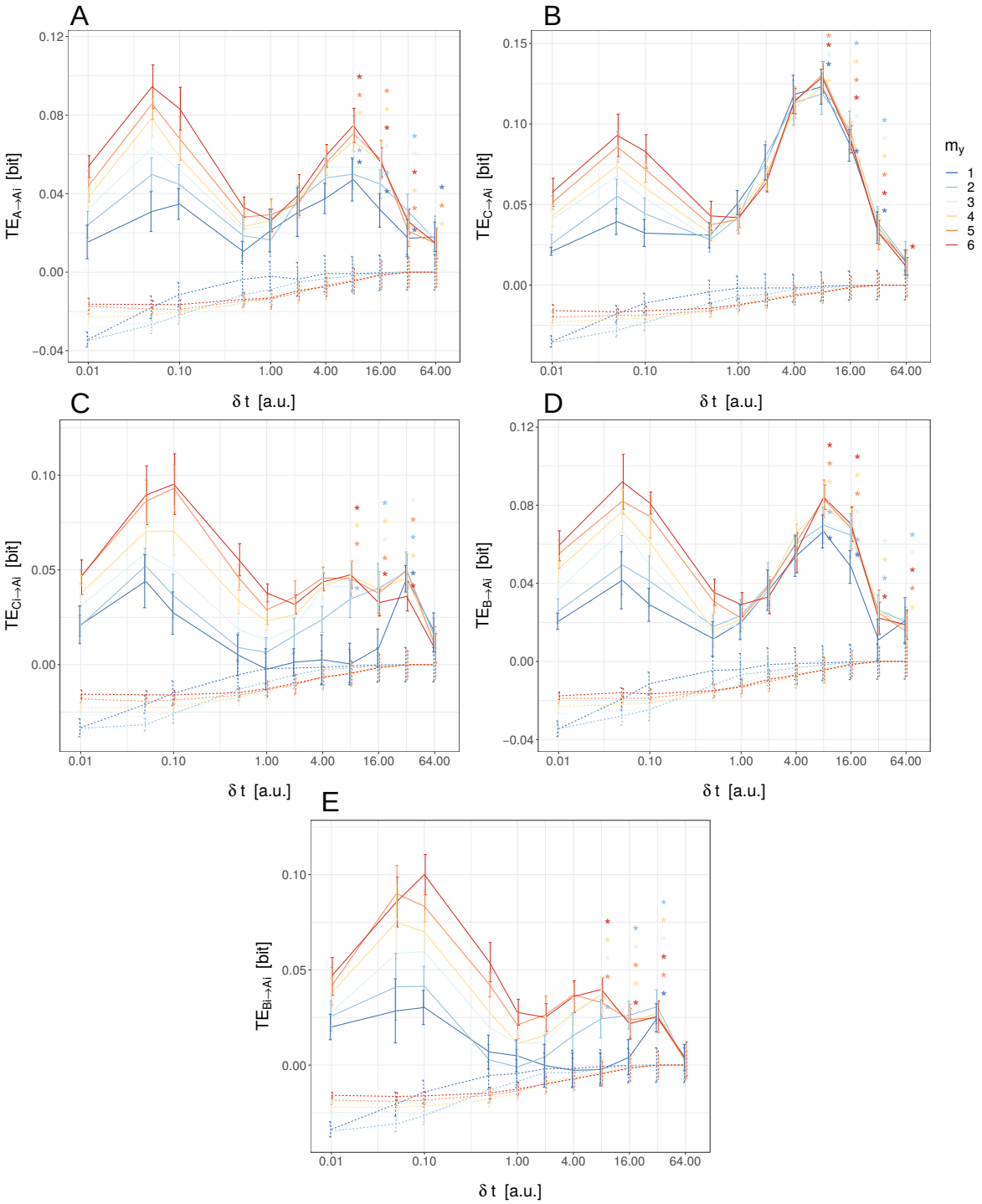

Figure A.6: δt-scan for model SB2.2. Unconditioned TE estimations in dependency of the sampling frequency δt and source history length m_y for the target species A_i are shown. Information transfer for different source-target constellations between the species of model SB2.2 are computed: **A:** $A \to A_i$, **B:** $C \to A_i$, **C:** $C_i \to A_i$, **D:** $B \to A_i$, **E:** $B_i \to A_i$. Among these, only panel **(B)** displays a truly connected coupling. Dashed lines indicate the respective surrogate estimations. Asterisks show significance of original estimates with difference to surrogates (one-sample Wilcoxon signed rank test with $p \leqslant 0.01$, Holm-Bonferroni corrected). Note that the significance level is only calculated for δt 8, 16, 32 and 64 [a.u.] with 20 original estimates per data point, where 1000 surrogate estimates correspond to one original estimate. Otherwise, for every original estimate (with a total of 10 original estimates), 10 surrogate estimation were performed, and no significance calculation was applied. Computations are realized with data length $N = 5000$ and $m_x = 2$ for the unvarying embedding dimension. The error bars indicate the standard deviation (sd). The x-axis is logarithmically scaled.

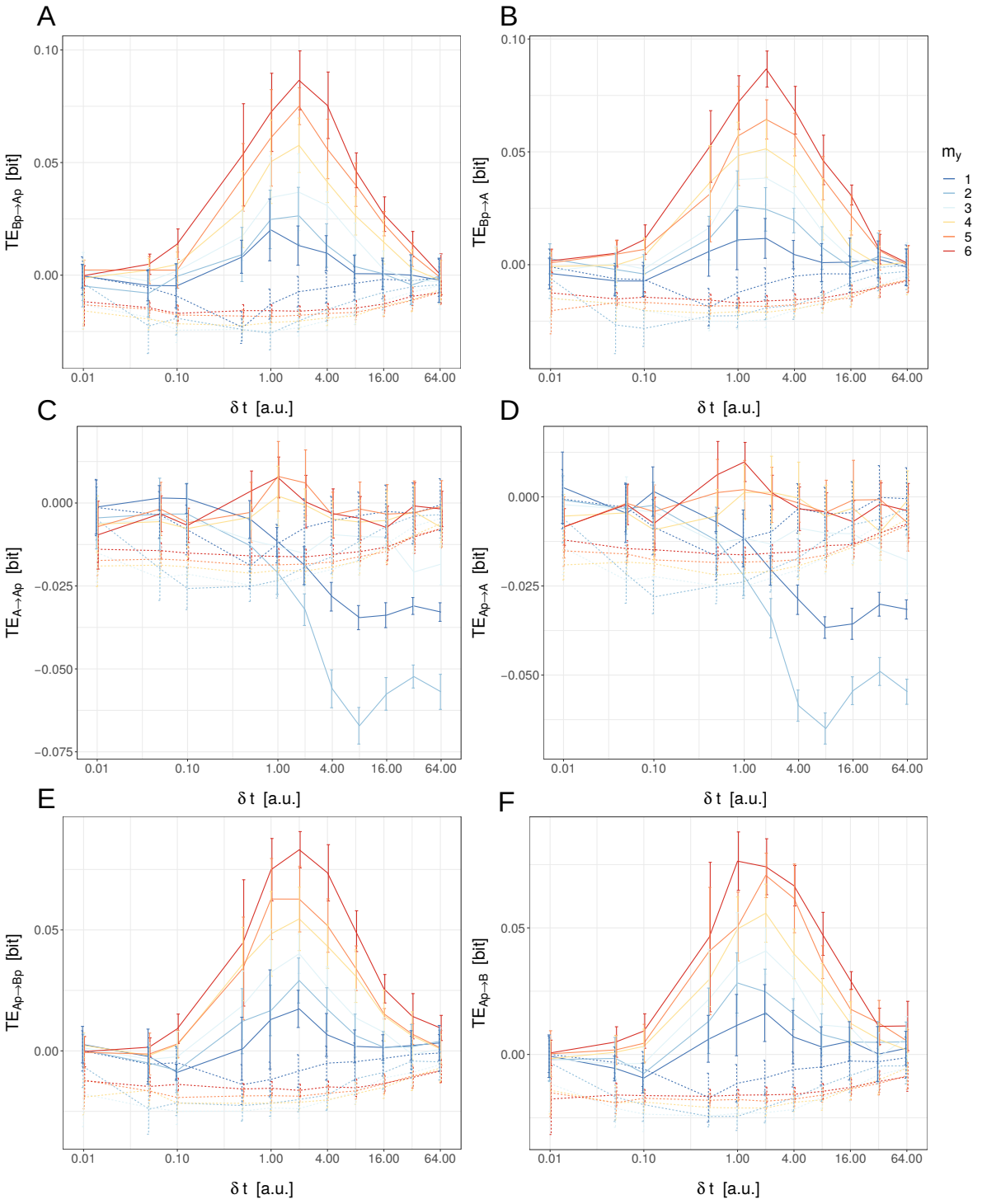

Figure A.7: δt-scan for model SM. Unconditioned TE estimations in dependency of the sampling frequency δt and source history length m_y are shown. Information transfer for different source-target constellations between the species of model SM are computed: **A:** $B_P \to A_P$, **B:** $B_P \to A$, **C:** $A \to A_P$, **D:** $A_P \to A$, **E:** $A_P \to B_P$, **F:** $A_P \to B$. All shown constellations are truly connected couplings. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original or 100 surrogate estimates respectively). Computations are realized with data length $N = 5000$ and $m_x = 2$ for the unvarying embedding dimension. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.

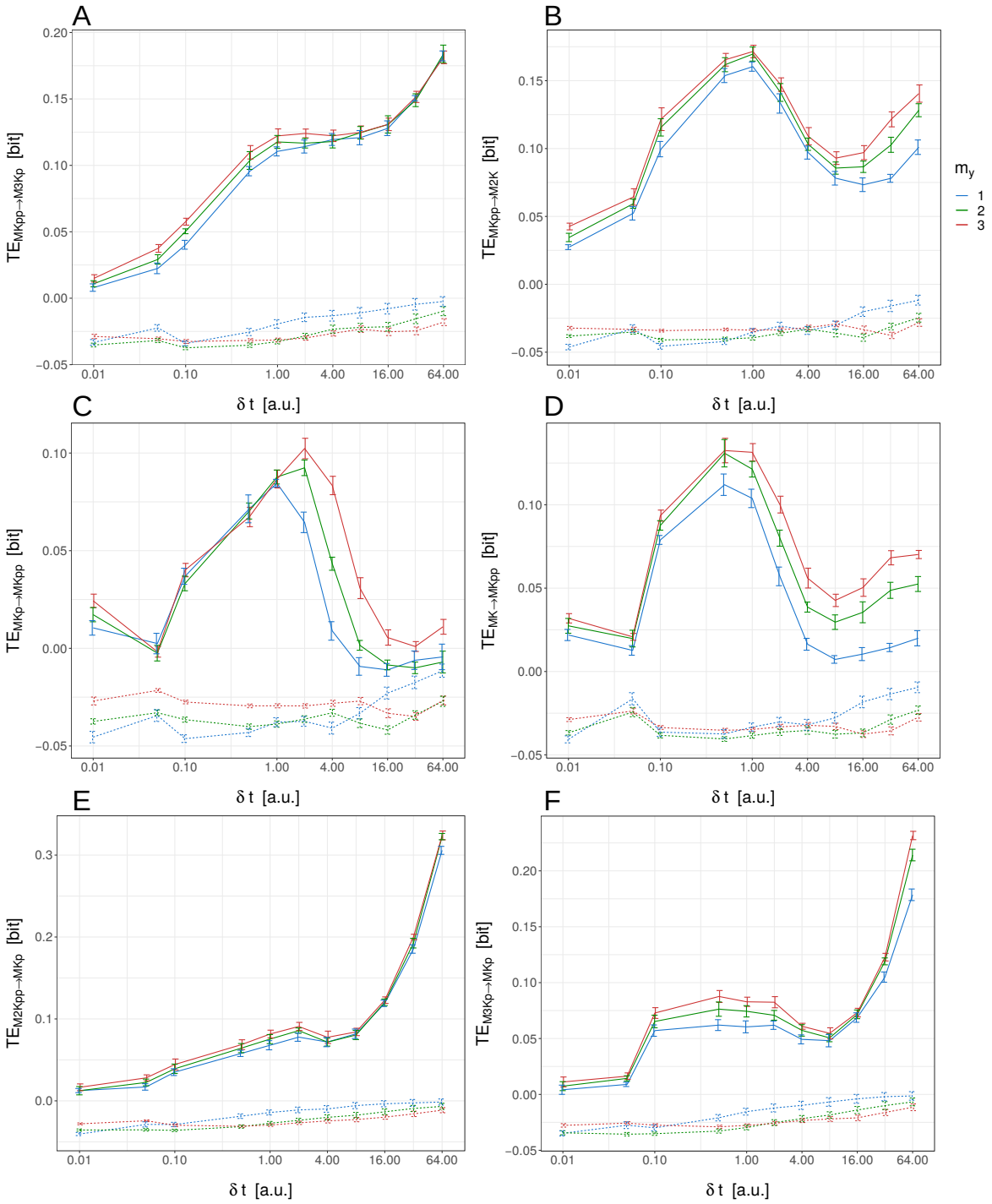

Figure A.8: δt-scan for model M1. Unconditioned TE estimations in dependency of the sampling frequency δt and source history length m_y are shown. Information transfer for different source-target constellations between the species of model M1 are computed: **A:** $MK_{PP} \to M3K_P$, **B:** $MK_{PP} \to M2K$, **C:** $MK_P \to MK_{PP}$, **D:** $M2K_{PP} \to MK_P$, **E:** $M2K_{PP} \to MK_P$, **F:** $M3K_P \to MK_P$. Panels on the left side exhibit truly connected couplings, whereas panels on the right side are unconnected constellations. The chosen source-target constellations correspond to possible information flow via different motifs: **A,B:** negative feedback, **C,D:** (de-)phosphorylation cycle, **E,F:** phosphorylation cascade. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original or 100 surrogate estimates respectively). Computations are realized with data length $N = 20000$ and $m_x = 2$ for the unvarying embedding dimension. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.

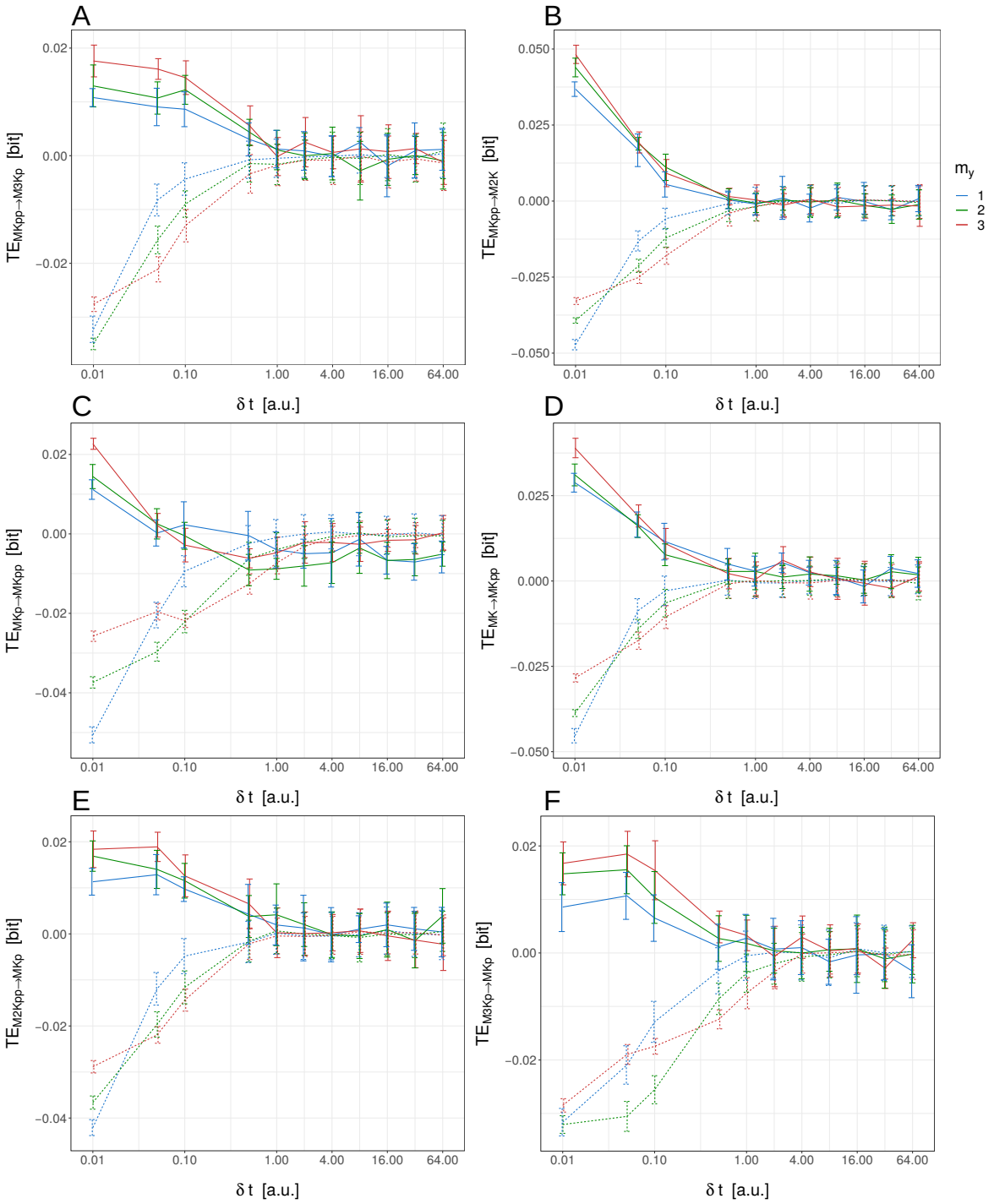

Figure A.9: δt-scan for model M2. Unconditioned TE estimations in dependency of the sampling frequency δt and source history length m_y are shown. Information transfer for different source-target constellations between the species of model M2 are computed: **A:** $MK_{PP}{\rightarrow}M3K_P$, **B:** $MK_{PP}{\rightarrow}M2K$, **C:** $MK_P{\rightarrow}MK_{PP}$, **D:** $M2K_{PP}{\rightarrow}MK_P$, **E:** $M2K_{PP}{\rightarrow}MK_P$, **F:** $M3K_P{\rightarrow}MK_P$. Panels on the left side exhibit truly connected couplings, whereas panels on the right side are unconnected constellations. The chosen source-target constellations correspond to possible information flow via different motifs: **A,B:** negative feedback, **C,D:** (de-)phosphorylation cycle, **E,F:** phosphorylation cascade. Dashed lines indicate the respective surrogate estimations. For every original estimate, 10 surrogate estimation were performed (in total, every point corresponds to 10 original or 100 surrogate estimates respectively). Computations are realized with data length $N = 20000$ and $m_x = 2$ for the unvarying embedding dimension. The error bars indicate the standard deviation (sd). Note that the x-axis is logarithmically scaled.