Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

VALIDATION WITHOUT DATA - FORMALIZING STYLIZED FACTS OF TIME SERIES

Pia Wilsdorf
Marian Zuska
Philipp Andelfinger
Adelinde M. Uhrmacher

Institute for Visual and Analytic Computing
University of Rostock
Albert-Einstein-Str. 22
Rostock, 18059, GERMANY

Florian Peters

Department of Economics
University of Rostock
Ulmenstraße 69
Rostock, 18057, GERMANY

ABSTRACT

A stylized fact is a simplified presentation of an empirical finding. When modeling and simulating complex systems and real data are sparse, stylized facts have become a key instrument for building trust in a model as they represent important requirements regarding the model's behavior. However, automatically validating stylized facts has remained limited as they are usually expressed in natural language. Therefore, we develop a formal language with a custom syntax and tailored predicates allowing modelers to unambiguously and succinctly describe important (temporal) characteristics of simulation traces or relationships between multiple traces via statistical tests. The proposed formal language is able to express numerous facts from the literature in different application domains, as well as to automatically check stylized facts. If stylized facts are defined at the beginning of a simulation study, formally expressing and checking them can streamline and guide the development of simulation models and their successive revisions.

1 INTRODUCTION

The validation of simulation models typically relies on the availability of data. Data are used to make expectations regarding the model's behavior explicit and to relate the model to the system of interest—often achieved by comparing time series produced by the model to time series observed in the real system. This is also referred to as operational validity (Sargent 2013) or output corroboration (Augusiak et al. 2014). However, in many areas such as economy, sociology, demography, or epidemiology, modelers lack high-quality data for validating simulation models (Augusiak et al. 2014). Moreover, when building these models, one often aims to capture abstract mechanisms (i.e., patterns or trends) of an observed phenomenon rather than reproducing one specific data set. Consequently, simulation studies increasingly use statements about the model behavior in the form of so-called stylized facts (Houy et al. 2015).

In economics, but also in the social sciences, demography, epidemiology, and ecology a stylized fact is defined as a simplified representation of an empirical finding, and thus a piece of generalized knowledge of a domain, enabling model validation or "pattern-oriented testing" without data from the real system (Madsen et al. 2021). In other complex systems, such as cyberphysical systems, cooperative multi-agent systems, and other systems of systems, although data might be available, validation is notoriously complex due to the need for specifying and checking requirements across multiple distributed system components. Utilizing stylized facts in these applications might be a cornerstone for making model validation feasible by focusing on the intended uses and central dynamics of the systems.

Typically, a set of stylized facts is defined for a simulation model. Referring to economics, stylized facts would be "With tax increase, the volatility of the GDP is likely to decrease" or "A lagged correlation

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

between firm indebtedness and credit defaults is to be expected" (Peters et al. 2022). Fulfilling a subset of these stylized facts will lend credibility to the simulation model and its explanatory power (Meyer 2019).

Stylized facts are expressed and communicated in natural language or a combination of mathematical formulas and natural language. This leaves the judgment of whether a model's behavior meets those stylized facts to the modeler's interpretation (Windrum et al. 2007), typically based on plots of the simulation output. This currently prevents stylized facts from becoming an integral part of a simulation study for which automated tool support exists, such as for automatic model validation.

Behavioral requirements or hypotheses regarding the expected simulation output can be formalized using temporal logics (e.g., Metric Interval Temporal Logic (MITL) (Maler and Nickovic 2004) or Signal Temporal Logic (STL) (Raman et al. 2014)). This enables the use of state-of-the-art (statistical) model checkers for automatic model validation and verification (Sebastio and Vandin 2013). Temporal logic, therefore, may be suited as a basis for unambiguously specifying stylized facts.

To exploit the potential of stylized facts for conducting simulation studies, in this paper, we develop a language that allows us to formalize stylized facts as a specific type of requirement on the time series produced by a simulation model (Ruscheinski et al. 2020). This formalization allows us to integrate them into the process of automatically validating models, and the overall modeling and simulation life cycle (Sargent 2013). The features of our language, including the various predicates supported, have been selected based on a literature study to support commonly used stylized facts of three different application domains. Those encompass 1) statistical tests for expressing behavioral properties of individual simulation traces (such as cyclic behavior), 2) statistical tests for defining properties that relate different simulation traces (e.g., by cross-correlation), and 3) metric interval temporal logic for expressing temporal relations within and across time series. The result is a formal language that shall enable modelers and domain experts (often unfamiliar with logic-based languages) to intuitively, unambiguously and succinctly define stylized facts. Further, we enable users to parameterize the language to include domain-specific interpretations (e.g., what it means for some variable to be "strongly" increasing), and to map language variables to concrete output variables of the simulation model. Our corresponding model checker takes time series data as input (i.e., trajectories or mean trajectories generated by a simulation model), and calculates a Boolean decision.

If stylized facts are defined at the beginning of a simulation study, the language and model checker can be used to systematically guide the development of the simulation model and its revisions (Ruscheinski et al. 2020). We will show and discuss the operationalization of stylized facts for the development of a macroeconomic model, during which a variety of stylized facts about the financial markets are specified and analyzed.

The paper is organized as follows. First, we discuss related work regarding the formal specification of behavioral properties and statistical model checking in Section 2. Next, we conduct a literature study to identify different kinds of stylized facts w.r.t. the behavior of time series in Section 3. This then presents the starting point for developing our approach for making stylized facts formally explicit for model validation in Section 4. Section 5 comprises a case study in finance, followed by discussion, conclusions and future work in Section 6.

2 RELATED WORK

Various languages and formats have been designed for specifying requirements on time series. In their work, Wainer et al. present an approach for the verification of DEVS models (Wainer et al. 2002). Requirements are specified by the modeler as input-output relationships between individual events, stating that when an input event is injected at a specific time, the expected output at a given time must be a specified value.

In contrast, temporal logics, such as STL (Raman et al. 2014) and MITL (Maler and Nickovic 2004), are used to describe the behavior of model variables over the course of a simulation. Variants of these temporal logics have been developed to define properties for systems with specific characteristics, such as spatially distributed components (Vissat et al. 2019).

Other languages allow combining various temporal logics and offer means for calculating probabilities and expected values based on simulation data, such as implemented in MultiVeStA (Sebastio and Vandin 2013). The hypothesis specification language FITS (Formulating, Integrating and Testing of Hypotheses in Computer Simulation) also includes means for assigning values to output variables of the stochastic simulations (e.g. the attributes of agents), information about the null hypothesis h_0 and an alternative hypothesis h_1 using formal logics, and finally parameters of the hypothesis test, such as a significance level (Lorig et al. 2017). The Biological Property Specification Language (BPSL) specifies properties as inequality constraints on the model output (Mitra et al. 2019). It thereby allows encoding quantitative or qualitative experimental data to be used in a model checking. It does not use temporal operators for the sake of readability, and thus only a subset of what is possible with temporal logics can be expressed.

In the realm of software requirements engineering, the trade-off between property specification in a formal logic-based versus natural language has been investigated by DeepSTL (He et al. 2022), which provides a first approach towards automatically deriving STL formulas from requirements specifications in the English language. DeepSTL also introduced additional temporal operators as syntactic sugar, to define falling or rising edges of signals. However, predicates for time series analysis beyond the STL operators have not yet been considered.

Similarly to DeepSTL, DeepOCL has explored natural language processing for generating constraints in the Object Constraint Language (OCL) (Yang et al. 2023). There is immense potential in investigating such an approach for facilitating the specification of stylized facts. However, handling ambiguity and domain-specific aspects when automatically translating the natural language statements remains a challenge (Diamantopoulos et al. 2017), and can be addressed using ontologies and domain-specific languages.

Our language caters specifically to the needs for specifying stylized facts of time series in two ways. First, it takes the temporal operators and bounds based on MITL but enriches those with the specification of statistical tests as part of the requirements. Those will be required for expressing properties of individual time series as well as the relation between time series. Second, the concise syntax of our language is inspired by natural language descriptions of stylized facts found in the various application domains (see next section).

3 STYLIZED FACTS IN THE LITERATURE

Our goal is to develop a language that allows users to express stylized facts as they are observed in several scientific domains. Stylized facts may also occur in various forms, such as descriptions of properties of time series, relationships between time series, or temporal conditions. To systematically identify a reasonably broad set of stylized facts to guide our initial language design, we conducted a literature search across three domains. We searched Google Scholar using the search string "stylized facts X" with X standing for one of the domains: finance, COVID-19, and demography. From the search results for each domain, we selected the three top-cited papers that met all of the following criteria: a)the paper is written in English, b) the core of the paper pertains to the domain, c) stylized facts are expressed explicitly, and d) the facts are related to the behavior of time series. The stylized facts from the nine papers selected according to these criteria were used as a starting point for developing the syntax of our formal language. Exemplary stylized facts that will be used in the following for illustration of the language are listed in Table 1. The examples show that the language for specifying stylized facts will require means for describing and comparing the shape of time series, aggregate values, auto- and cross-correlations, cyclic behaviors, as well as temporal relationships. The following will address these requirements.

4 LANGUAGE DESIGN

In our language each stylized fact consists of one or more expressions, each terminated by a semicolon. Expressions can be composed by conjunctions, and nested. Of course, complex stylized facts can also be split into subfacts to improve their readability. Since during model checking, subfacts can be analyzed

Table 1: Stylized facts found in the literature related to different application domains.

Domain	Publication	Stylized Fact
Finance	Hommes (2002)	Asset returns have fat tails.
		Autocorrelations of squared returns are significantly positive, even at high-order lags.
	Cont (2001)	Autocorrelation of absolute returns decays as function of time lag.
		Most measures of volatility of asset are negatively correlated with returns of asset.
		Coarse grained measures of volatility predict fine-scale volatility better than the other way around.
	Fiorito (1997)	Government deficits adjust quickly to business cycle and are strongly countercyclical.
COVID-19	Knittel and Ozaltun (2020)	[T]here is evidence that all modes of commutes, other than biking, are associated with higher death rates relative to telecommuting.
Demography	Cervellati and Sunde (2011)	The demography literature has proposed several criteria to identify whether a country has reached the critical turning point of the demographic transition: a) Life expectancy at birth exceeds 50 years; b) Fertility or the crude birth rate has exhibited a sustained decline; c) The crude birth rate has fallen below the threshold of 30/1000.
	Williamson (1998)	[I]t appears that changing demographic conditions might have accounted for even higher proportions of the trend acceleration from low growth rates prior to 1970 to the 6.1% per annum rate afterward.

separately, this may also improve the model checking efficiency. For instance, the stylized fact by Cervellati and Sunde (2011) (Table 1) may be broken down into an expression about what it means for a society to be “pre-transitional” (characterized by high birth and high death rates), and an expression for “post-transitional” (low birth, and low death rates), and the various conditions of these two stages of a society. If all subfacts evaluate to true, the overall fact is also fulfilled.

In the following, we present the various features of our language using the examples introduced in Table 1. We start with simple expressions about the properties of individual time series. Then we continue with expressions for describing the relation between time series, and finally, those expressions are interlinked via MITL. The grammar of the language can be found at https://git.informatik.uni-rostock.de/mosi/stylized-facts.

4.1 Time Series and their Properties

We target simulation models that produce time series as output. If simulations are stochastic, we assume that aggregate trajectories are provided and thus the handling of replications and statistical model checking is not supported.

Expressions can comprise descriptions of the columns of a tabular data set and descriptions of scalars, i.e., aggregate values on the columns of the output data. Note that column names (output variables) can be arbitrary strings but that for the sake of readability and unambiguity, they have to be indicated with a leading underscore.

An expression may refer to a single time series (atomic column) or multiple columns combined by arithmetic operators (+, −, $\star$, or /). In addition, the language provides pre-defined functions on columns, such as cumulated(), squared(), or absolute(). Functions are also available for calculating summary statistics, including the mean, standard deviation, variance, and the first element of the time series. The return values of these functions then may be compared to other scalars. The following expression states that the growth rate of the population is required to be greater than six percent (Williamson 1998):

```
growthrate(_population) > 0.06;
```

In addition, time series can be assigned attributes, describing their shape as (monotonically) increasing, falling, power law, persistent, stationary, heavy-tailed, light-tailed, or flat. For instance, the time series of the returns may be described as fat-tailed (Hommes 2002):

```
_returns are fat-tailed;
```

Additional predicates provide access to statistical tests, for example, to test for autocorrelation with a specific lag, or for kurtosis (i.e., the skewness of the time series values over time). The results of these tests may then be characterized as being positive, negative, or close to zero. Furthermore, these expressions can be associated with a strength, such as strong, medium, or small. As one example, the autocorrelation of the squared returns may be described as being strong positive if the lag equals 2 (Hommes 2002):

```
AC(squared(_returns), lag=2) is strong positive;
```

4.2 Relation between Time Series

One way relating two time series is through comparisons of aggregate values (as seen above). Moreover, simple point-wise comparisons of the time series are possible. For instance, according to Knittel and Ozaltun (2020), during the COVID-19 pandemic, the death rate of people working from home (telecommuting) was lower than the death rate of people commuting by car or public transit. The death rate associated with biking, however, was lower than the one for telecommuting. This can be expressed as three separate facts:

```
_death_rate_telecommuting < _death_rate_car;
_death_rate_telecommuting < _death_rate_public;
_death_rate_telecommuting > _death_rate_bike;
```

Another important criterion for relating time series is cyclicality. Especially in the context of finance it is often used to describe the deviation from business trends. A variable is considered to behave procyclically if it is positively correlated with the state of the economy (business cycle), whereas variables with negative effect on the economy are considered to be countercyclical. A cyclic variable can be leading or lagging, i.e., be cyclic with a positive or negative shift in the x-axis (by a number of time periods). The publication by Fiorito (1997) describes the government deficits to be counter-cyclically lagging behind the business cycles (given by the real GDP), which is modeled as:

```
_government_deficits are countercyclically lagging _real_gdp;
```

Furthermore, cross-correlations—as a measure of similarity between two columns—can be specified explicitly. For instance, in the first example below, the trading volume and volatility are required to be positive (Cont 2001). Moreover, the cross-correlations of various time series may be put into relation with each other. Here also a lag may be specified, which represents the allowed displacement of the two time series in number of time periods, in this case 2 periods:

```
CC(_trading_volume, _volatility) is positive;
CC(_coarse_volatility, _fine_volatility, lag=2) > CC(_fine_volatility,
    _coarse_volatility, lag=2);
```

4.3 Temporal Relations within and across Time Series

Finally, the expressions shown above can be combined to more complex expressions via temporal operators. Based on our literature review, we identified, two unary temporal operators, i.e., `Always` and `Eventually`, to be of relevance for our language. Always expresses that the constraint has to be fulfilled at each point in the time series, whereas Eventually means that the fact only needs to be true at some point in the time series—it may even become false again after that. The semantics of these two operators we defined according to common temporal operators often denoted as G or $\Box$ (globally) and F or $\Diamond$ (finally) (Maler and Nickovic 2004).

The fact by Cervellati and Sunde (2011) lists the conditions for a country to be regarded as post-transitional. One of those conditions can be expressed as follows by combining the use of Eventually and Always:

```
Eventually Always _life_expectancy > 50;
```

Binary temporal operators (syntax: a operator b) of our language include `Until`, `Precedes`, `Follows`, and `After`. Again, we identified those operators as possibly relevant based on our literature study. The `Until` operator has the same semantics as the Until operator commonly used in temporal logic (Maler and Nickovic 2004): with `a Until b`, we expect `b` to become true at some point (and before this point `a` has to be true). Using the expression `a Precedes b`, we express that `a` may not occur after `b`, and if `b` is encountered at some point, `a` must have been true before this point. `Follows` is interpreted similarly to the Release operator of temporal logic: `a Follows b` implies that `b` has to be true until and including the point where `a` first becomes true. `a After b` denotes that `a` may not occur before `b` and if `b` is encountered at some point, `a` will eventually become true afterward. Obviously, we could have interpreted these operators also differently. To communicate the meaning of operators to the modelers, make them accessible, and allow modelers to adjust interpretations to their own perception is subject of future research.

Since we are building on MITL, expressions can be associated with some temporal bounds or durations, as well as units (e.g., a specific number of days or years) for which the expression must be fulfilled. To allow nesting of those (temporal) logic expressions, the operators `and`, `or`, and `not` can be used. The following expression combines these features to describe the condition for being a post-transitional country, thereby relating the time series of life expectancy, crude birth rate (CBR), and fertility, see the stylized fact by Cervellati and Sunde (2011). In this fact, the condition `b` ("fertility is falling...") is expected to become true within 5 to 20 time units after condition `a`:

```
(_life_expectancy > 50 and _CBR < 0.03) after (_fertility is falling for 5 years or
    _CBR is falling for 5 years) within 5 to 20 years;
```

Additionally, we provide the quantifiers `Forall` and `Exists` to allow specifying an expression over all lags in a certain range. Above we have seen the stylized fact that the autocorrelation of squared returns is strong positive for lag 2 (Hommes 2002). This property we now want to be satisfied for higher lags as well:

```
Forall <lag> in {2,5,7}: AC(squared(_returns), lag=<lag>) is strong positive;
```

Similarly, ranked quantifier expressions are defined together with the phrase "the smaller/larger" with respect to the value in a list of lags, or "the smaller/larger" with respect to some property, such as the autocorrelation of absolute returns (Cont 2001):

```
The larger <lag> in {1,2,3,4} the smaller AC(absolute(_returns), lag=<lag>);
```

4.4 Parametrization of the Language

As a consequence of our language's goal of bridging the gap between stylized facts expressed in natural language and formal logics, there is the need to apply a precise definition of notions such as "increasing" or even "strongly increasing". The thresholds beyond which such properties are considered to hold true can be domain-specific. To account for this, language constructs such as "increasing" are parametrized with a corresponding threshold. Our vision is for researchers from a specific domain to agree on default parameters that are assumed when no additional information is given as part of any publication that lists stylized facts a simulation model accords with. If a particular stylized fact requires deviating parameters to satisfy a specified property, the authors would specify and justify this choice. This approach provides the flexibility needed to cater to different domains, while unambiguously communicating the assumptions under which facts hold true and how the facts can be checked independently. Ontologies specialized for a specific application (or even user) may assist in disambiguating the different interpretations.

5 CASE STUDY

In the following, we showcase the use of our language to express stylized facts and automatically check whether they hold true for result produced by an agent-based simulation. The facts assessed in this case study were not considered during the design of the language but could all be expressed using the language constructs described in Section 4.

5.1 Model and Data

In Peters et al. (2022), an agent-based model has been developed to analyze the impact of monetary reforms on macro-financial stability, or in general on the dynamics of financial crises. To calibrate and validate the model, a method common in economics was adopted, i.e., to check whether the simulation model can reproduce a range of macroeconomic and microeconomic stylized facts (Dosi et al. 2017). Thirteen stylized facts were tested. For this, micro-founded simulation results were aggregated to describe properties at macro level, such as the Gross Domestic Product (GDP). Then, trajectory visualizations or summary statistics in tables were displayed to support the argumentation of whether a stylized fact applies. Thereby, the relation between stylized facts, argumentation, and simulation results remained implicit. In contrast, in the following, we make this relation explicit and unambiguous.

We, therefore, have implemented the presented language and model checker using Python and ANTLR (Parr 2013) and made the code available at https://git.informatik.uni-rostock.de/mosi/stylized-facts. The simulation model of this case study, its output data (time series), and additional R scripts required for preprocessing are available at https://github.com/fhaegner/Mak-h-ro. To collect the time series data, the simulation model was run for 20 stochastic replications and 800 time periods (one period refers to approx. one quarter in the real economy). Note that the model is not designed for forecasts or to capture economic growth or technology change, and therefore is not restricted to a particular finite simulation time horizon. What is of interest is the interaction of the defined theoretical framework. Preprocessing on the model output was done to cut off a warm-up phase of 200 periods, aggregate trajectories (i.e., the means per period) are calculated, and the time series were log transformed, except for inflation and stock change, which exhibit negative values. Finally, a band-pass filter was applied to extract the stationary "business cycle component" within a particular band of frequencies, thereby "detrending" the data (Baxter and King 1999).

5.2 Validation using Stylized Facts

The simulation model is built with the objective of simulating behavior in financial crises. Thus, it must meet the essential criteria of a business and credit cycle. Hence, straightforward and salient stylized facts are defined at the beginning of the simulation study based on literature and empirical data. The corresponding publication of the model lists thirteen stylized facts (Peters et al. 2022). Out of those we selected the SFs 1–4 to showcase our language and model checker.

During model building, these facts are constantly checked as they form important requirements regarding the model's behavior. For each major model revision, a subset of stylized facts holds true while others are still unsatisfied. They thereby guide the successive model refinements, finally leading to the validated (and published) model version. In the following, we formally express the stylized facts using our language, and validate them based on output data of the final (published) model version (Peters et al. 2022).

The first stylized fact (**SF1**) refers to the behavior at business cycle frequencies related to the GDP, firm investment and consumption time series. The volatility (defined as the standard deviation of the time series) of investment is expected to be greater than the GDP fluctuations (Stock and Watson 1999). In contrast, consumption is expected to be less volatile than the GDP. In our language this stylized fact reads:

```
standard deviation(_real_investment) > standard deviation(_gdp_real) and standard
    deviation(_consumption) < standard deviation(_gdp_real);
```

Our automatic model checker finds this fact to be true for the given model, which accords with the plot shown in Figure 1.

According to Wright (2005), the recession duration is defined as the length of periods in which the real GDP percentage change is less than zero. **SF2** requires the duration of the recession periods to be falling over time and, in addition, to be heavy-tailed. Since the recession period length is one of the model's output

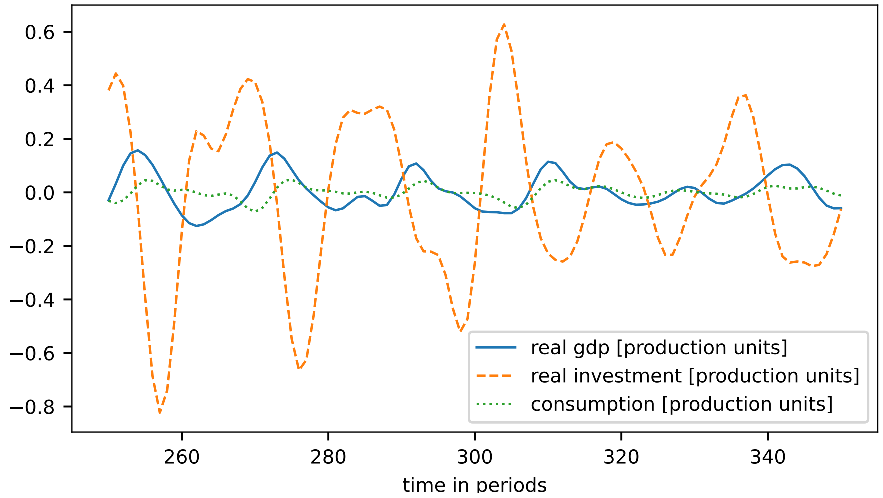

Figure 1: Plot of 100 time periods of the simulation data regarding SF1. The fluctuations of the real GDP are smaller than the fluctuations in real investment but larger than the fluctuations in consumption. Thus, SF1 is satisfied.

variables, the stylized fact can be formalized as follows and be evaluated as true by our model checker, as shown in Figure 2:

```
_recession_period_length is falling and _recession_period_length is heavy-tailed;
```

This stylized fact is important for the model building process to check whether the model captures simple empirical observations of crisis dynamics within the real economy. Empirical observations led to the conclusion that severe crises with a long duration of shrinking GDP (e.g., the financial crisis of 2007/08 with a duration of up to 6 quarters) (Aiginger 2009) appear to be infrequent or as an anomaly while short recessions (i.e., 2 quarter) are interpreted as an artifact of a usual business cycle (Wright 2005).

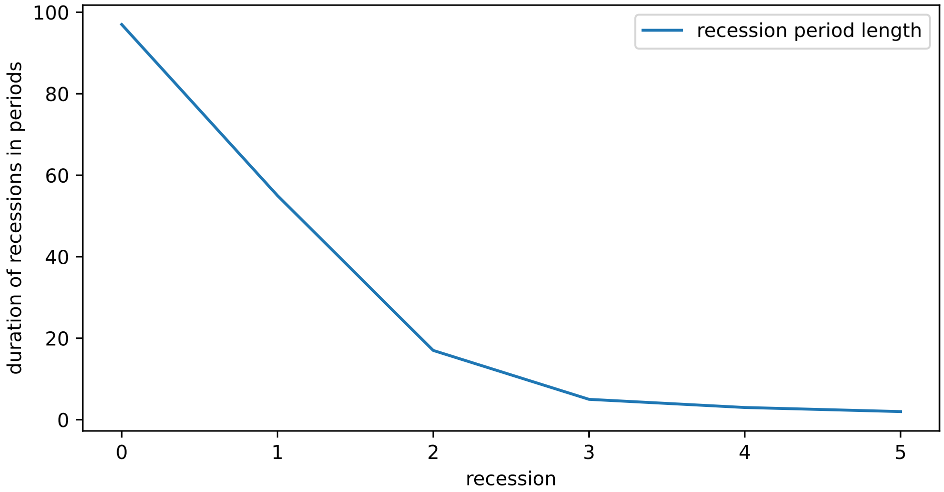

Figure 2: Plot of the recession period length, exhibiting a heavy-tailed curve.

The next stylized fact, **SF3**, refers to the relation of various macro variables. Investment, change in inventories (stock), inflation and prices are expected to be procyclical, which is in line with observations made by Wälde and Woitek (2004) and (Apergis 1996) (for prices). This translates to the following expression in our language, with the chosen lag values being based on domain knowledge (Stock and Watson 1999):

```
_stock_change is procyclically lagging _gdp_real by 1 lag and _inflation is
    procyclically lagging _gdp_real by 3 lags and _price are procyclically lagging
    _gdp_real by 1 lag and _unemployment is countercyclical to _gdp_real;
```

For model checking the cyclicality, in our case study, we parametrize the correlation thresholds for a time series to be considered as procyclical (correlation >0.5) or countercyclical (correlation <-0.5). Checking this fact with these thresholds on the simulation results confirms the findings of Wälde and Woitek (2004). In Figure 3, the involved time series are visualized and the described procyclical and counter-cyclical behaviors can be clearly identified.

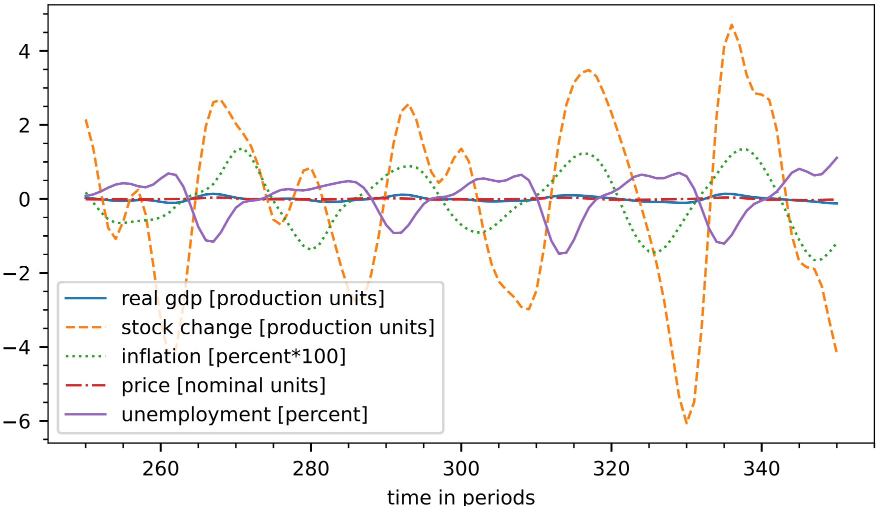

Figure 3: Plot of 100 time periods of the simulation data regarding SF3. Procyclical behaviors of investment, stock change, inflation as well as prices, and counter-cyclical behavior of unemployment with respect to real GDP can be seen.

SF4 looks closer at the investments of firms in research and development. Empirical observations suggest a nearly congruent behavior of firm investment and real GDP, i.e., procyclical behavior with lag 0 (Wälde and Woitek 2004). The model by Peters et al., however, is designed to capture financial crises behavior, for which the investment leads real GDP with a lag between 3 and 7 time periods. Thus, for this case study, the stylized fact is modified to:

```
_investment is procyclically leading _gdp_real within 3 to 7 lag;
```

Automatically checking the modified SF4 reveals that it is indeed satisfied by the model, whereas testing the original stylized fact by Wälde and Woitek (2004) with lag 0 is not successful (see Figure 4). Thus, the simulation model behaves as desired.

6 CONCLUSIONS AND OUTLOOK

Based on the requirements and research gaps identified in the literature review, we developed a language that is specialized for expressing stylized facts of time series. In particular, various statistical tests on time series needed to be incorporated with temporal-logic properties inspired by metric interval temporal logic. The objective was to use a controlled vocabulary that is close to natural language to be intelligible by domain modelers. As proof of concept, we applied the language in an economic simulation study, and checked whether the simulation model conforms to the defined stylized facts.

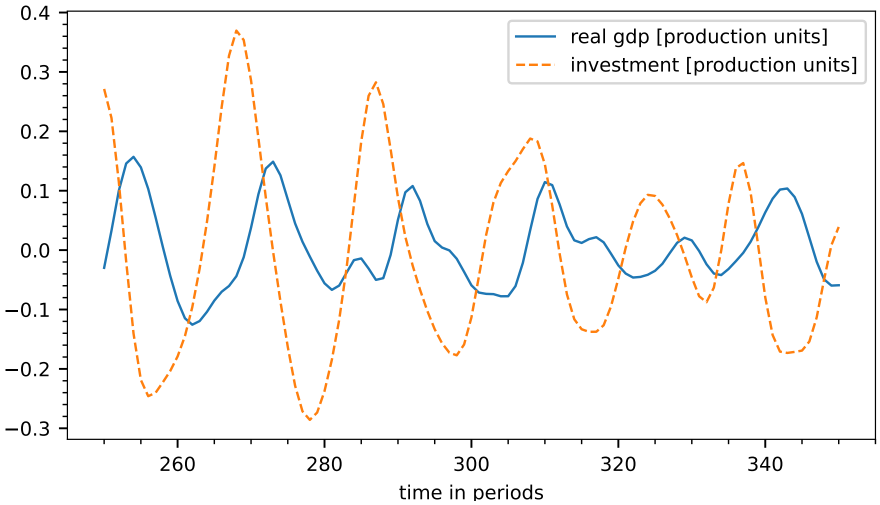

Figure 4: Plot of 100 time periods of the simulation data regarding SF4. It shows investment leading real GDP by 3–7 periods.

The use of a set of stylized facts increases the credibility of and trust in simulation models in domains where data are scarce. They allow specifying more abstract patterns and focusing on theoretical implications based on empirical findings. Our case study aims to conduct what-if analyses of theoretical economic implications. In combination with the "what-if" scenarios, a validation based on general empirical patterns (stylized facts) appears more promising than validation on specific historical events or periods as it avoids overfitting.

Formalizing stylized facts makes them accessible for communication and exchange, as they can explicitly and unambiguously be shared as part of scientific publications or model documentation. Also, as with any formal language, it may serve as a thinking tool by prompting the modeler to specify properties unambiguously and to check different kinds of (non-obvious) properties with the model. In addition, formalization makes it accessible for automatic interpretation. From the modelers' perspective, an automatic validation tool for stylized facts is beneficial beyond the final validation of the model. If stylized facts are specified early in the simulation study, automatic checks can guide and streamline the model development process (Ruscheinski et al. 2020) and help in creating simulation models that conform to important empirical findings. It should be noted that formally specified stylized facts cannot be expected to cover all relevant aspects of a simulation model, thus validation by stylized facts is not a substitute for other validation methods, but particularly for domains where data are scarce, a valuable addition to the portfolio of validation methods (Leye et al. 2009).

For being a valuable addition, an easily accessible and sufficiently expressive language referring to syntax and semantics is paramount. An intelligible syntax is needed for the modeler to adopt the tool as a thinking tool. Therefore its syntax should feel natural to the domain experts; it should allow them to express what is needed in a compact and succinct manner. Currently, the language provides a range of features that allow expressing common stylized facts, but it is by no means complete. In future work, the language may be extended to suit more specific needs of time series analysis or particular application domains. For instance, supporting the empirical characterization of probability distributions on time series would allow us to express the remaining stylized facts from the simulation study of Peters et al. (2022), for instance, SF8: “the fiscal costs of banking crises-to-GDP ratio has a fat-tailed distribution”. Adding support for the required statistical tests to our language and model checker is possible with marginal development effort. We are also considering features to support stylized facts requiring analyses across simulation replications. This may include determining confidence intervals over stochastic simulations. However, the integration of relevant additional features has to be done in light of its impact on the compactness of the language in order to maintain its accessibility for modelers from various domains.

So far we only discussed the challenge of syntax; however, to easily access the semantics of the language is as important for the modeler, so that the modeler knows what is meant by specifying expressions like `a after b`. Does this mean we assume `a` and `b` to be true at some point in the trajectories and that `b` follows directly `a`? What does strongly positively correlated mean? Consequently, means are required to involve the modeler (and possibly modeling communities) in the design of such languages for stylized facts, referring to syntax and supported features as well as their interpretations, i.e., their semantics. Thereby, the languages will necessarily evolve with the need of the applications and the communities.

The semantics of a stylized fact and the validation result may also be impacted by the preprocessing of the simulation output. Extensions of the language, therefore, may allow making the various preprocessing steps (e.g., removing the initial transient or applying filters) and their parametrizations explicit. Alternatively, the context in which stylized facts are evaluated may be stored with the provenance of the simulation study (Ruscheinski and Uhrmacher 2017). In particular, provenance graphs provide a standardized form that allows documenting which fact was checked and satisfied based on what conditions and simulation data, and how this data was collected. In addition, there are different ways of interpreting the validation result. So far, the satisfaction of a stylized fact was a binary decision: all subfacts had to be fulfilled for the entire statement to become satisfied. Extensions of the language may pursue a probabilistic approach by allowing to specify a measure for the degree of fulfillment.

REFERENCES

Aiginger, K. 2009. “A Comparison of the Current Crisis with the Great Depression as Regards their Depth and the Policy Responses”. Working Paper 354, WIFO. https://www.oecd.org/economy/growth/43734645.pdf.

Apergis, N. 1996. “The Cyclical Behavior of Prices: Evidence from Seven Developing Countries”. *The Developing Economies* 34(2):204–211.

Augusiak, J., P. J. Van den Brink, and V. Grimm. 2014. “Merging Validation and Evaluation of Ecological Models to ‘Evaludation’: A Review of Terminology and a Practical Approach”. *Ecological Modelling* 280:117–128.

Baxter, M., and R. G. King. 1999. “Measuring Business Cycles: Approximate Band-pass Filters for Economic Time Series”. *Review of Economics and Statistics* 81(4):575–593.

Cervellati, M., and U. Sunde. 2011. “Life Expectancy and Economic Growth: the Role of the Demographic Transition”. *Journal of Economic Growth* 16:99–133.

Cont, R. 2001. “Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues”. *Quantitative Finance* 1(2):223.

Diamantopoulos, T., M. Roth, A. Symeonidis, and E. Klein. 2017. “Software Requirements as an Application Domain for Natural Language Processing”. *Language Resources and Evaluation* 51(2):495–524.

Dosi, G., M. Napoletano, A. Roventini, and T. Treibich. 2017. “Micro and Macro Policies in the Keynes+ Schumpeter Evolutionary Models”. *Journal of Evolutionary Economics* 27:63–90.

Fiorito, M. R. 1997. “Stylized Facts of Government Finance in the G-7”. Working Paper 1997/142, International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2016/12/30/Stylized-Facts-of-Government-Finance-in-the-G-7-2368.

He, J., E. Bartocci, D. Ničković, H. Isakovic, and R. Grosu. 2022. “DeepSTL: From English Requirements to Signal Temporal Logic”. In *Proceedings of the 44th International Conference on Software Engineering*, ICSE ’22, 610–622. New York, NY, USA: Association for Computing Machinery.

Hommes, C. H. 2002. “Modeling the Stylized Facts in Finance through Simple Nonlinear Adaptive Systems”. *Proceedings of the National Academy of Sciences* 99(suppl_3):7221–7228.

Houy, C., P. Fettke, and P. Loos. 2015. “Stylized Facts as an Instrument for Literature Review and Cumulative Information Systems Research”. *Communications of the Association for Information Systems* 37(1):10.

Knittel, C. R., and B. Ozaltun. 2020. “What Does and Does Not Correlate with COVID-19 Death Rates”. Working Paper 27391, National Bureau of Economic Research. http://www.nber.org/papers/w27391.

Leye, S., J. Himmelspach, and A. M. Uhrmacher. 2009. “A Discussion on Experimental Model Validation”. In *2009 11th International Conference on Computer Modelling and Simulation*, 161–167. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lorig, F., C. A. Becker, and I. J. Timm. 2017. “Formal Specification of Hypotheses for Assisting Computer Simulation Studies”. In *Proceedings of the Symposium on Theory of Modeling & Simulation*, 1–12. San Diego, CA: Society for Computer Simulation International.

Madsen, J. K., R. Bailey, E. Carrella, and P. Koralus. 2021. “From Reactive Towards Anticipatory Fishing Agents”. *Journal of Simulation* 15(1-2):23–37.

Maler, O., and D. Nickovic. 2004. “Monitoring Temporal Properties of Continuous Signals”. In *International Symposium on Formal Techniques in Real-Time and Fault-Tolerant Systems*, 152–166. Berlin, Heidelberg: Springer.

Meyer, M. 2019. “How to Use and Derive Stylized Facts for Validating Simulation Models”. *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*:383–403.

Mitra, E. D., R. Suderman, J. Colvin, A. Ionkov, A. Hu, H. M. Sauro, R. G. Posner, and W. S. Hlavacek. 2019. “PyBioNetFit and the Biological Property Specification Language”. *iScience* 19:1012–1036.

Parr, T. 2013. “The Definitive ANTLR 4 Reference”. *The Definitive ANTLR 4 Reference*:1–326.

Peters, F., D. Neuberger, O. Reinhardt, and A. Uhrmacher. 2022. “A Basic Macroeconomic Agent-based Model for Analyzing Monetary Regime Shifts”. *PLOS ONE* 17(12):1–39.

Raman, V., A. Donzé, M. Maasoumy, R. M. Murray, A. Sangiovanni-Vincentelli, and S. A. Seshia. 2014. “Model Predictive Control with Signal Temporal Logic Specifications”. In *53rd IEEE Conference on Decision and Control*, 81–87. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ruscheinski, A., and A. Uhrmacher. 2017. “Provenance in Modeling and Simulation Studies—Bridging Gaps”. In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D’Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 872–883. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ruscheinski, A., T. Warnke, and A. M. Uhrmacher. 2020. “Artifact-Based Workflows for Supporting Simulation Studies”. *IEEE Transactions on Knowledge and Data Engineering* 32(6):1064–1078.

Sargent, R. G. 2013. “Verification and Validation of Simulation Models”. *Journal of Simulation* 7(1):12–24.

Sebastio, S., and A. Vandin. 2013. “MultiVeStA: Statistical Model Checking for Discrete Event Simulators”. Working Paper, IMT Institute for Advanced Studies Lucca. http://eprints.imtlucca.it/id/eprint/1798.

Stock, J. H., and M. W. Watson. 1999. “Business Cycle Fluctuations in US Macroeconomic Time Series”. *Handbook of Macroeconomics* 1:3–64.

Vissat, L. L., M. Loreti, L. Nenzi, J. Hillston, and G. Marion. 2019. “Analysis of Spatio-Temporal Properties of Stochastic Systems Using TSTL”. *ACM Transactions on Modeling and Computer Simulation* 29(4).

Wainer, G., L. Morihama, V. Passuello et al. 2002. “Automatic Verification of DEVS Models”. In *Proceedings of the 2002 Spring Simulation Interoperability Workshop*. Orlando, FL: Simulation Interoperability Standards Organization (SISO).

Wälde, K., and U. Woitek. 2004. “R&D Expenditure in G7 Countries and the Implications for Endogenous Fluctuations and Growth”. *Economics Letters* 82(1):91–97.

Williamson, J. G. 1998. “Growth, Distribution, and Demography: Some Lessons from History”. *Explorations in Economic History* 35(3):241–271.

Windrum, P., G. Fagiolo, and A. Moneta. 2007. “Empirical Validation of Agent-Based Models: Alternatives and Prospects”. *Journal of Artificial Societies and Social Simulation* 10(2):8.

Wright, I. 2005. “The duration of recessions follows an exponential not a power law”. *Physica A: Statistical Mechanics and its Applications* 345(3-4):608–610.

Yang, Y., Y. Liu, T. Bao, W. Wang, N. Niu, and Y. Yin. 2023. “DeepOCL: A Deep neural Network for Object Constraint Language Generation from Unrestricted Nature Language”. *CAAI Transactions on Intelligence Technology*:1–14.

AUTHOR BIOGRAPHIES

PIA WILSDORF is a Ph.D. candidate in the Modeling and Simulation group at the University of Rostock. Her e-mail address is pia.wilsdorf@uni-rostock.de.

MARIAN ZUSKA is a Master’s student of Computer Science and a student assistant in the Modeling and Simulation group at the University of Rostock. His e-mail address is marian.zuska@uni-rostock.de.

PHILIPP ANDELFINGER is a Postdoctoral researcher in the Modeling and Simulation group at the University of Rostock. His e-mail address is philipp.andelfinger@uni-rostock.de.

FLORIAN PETERS is a Ph.D. candidate in the Money and Credit group at the University of Rostock. His e-mail address is florian.peters@uni-rostock.de.

ADELINDE M. UHRMACHER is professor at the Institute for Visual and Analytic Computing, University of Rostock, and head of the Modeling and Simulation group. Her e-mail address is adelinde.uhrmacher@uni-rostock.de.