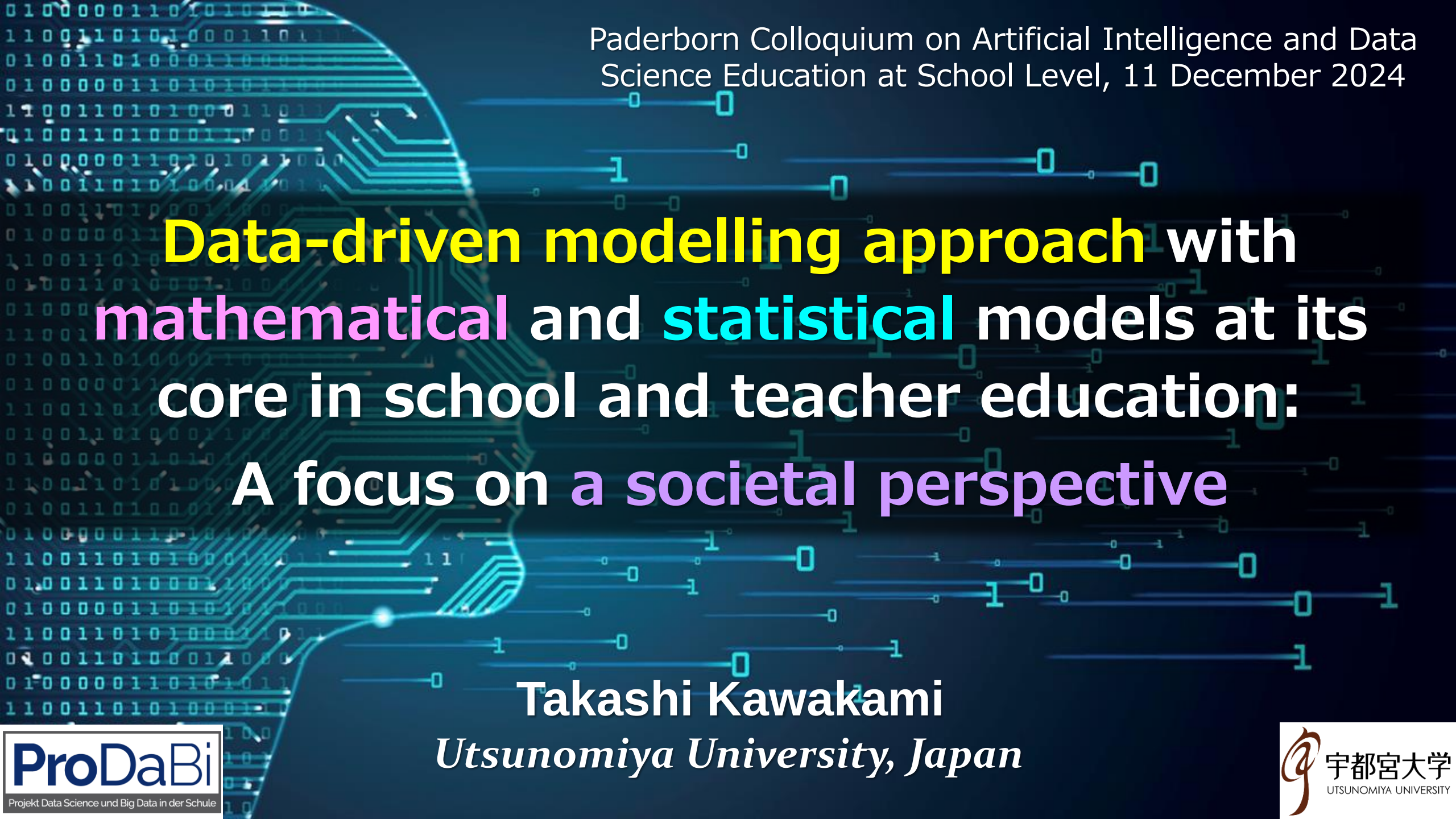


Paderborn Colloquium on Artificial Intelligence and Data Science Education at School Level, 11 December 2024

Data-driven modelling approach with mathematical and statistical models at its core in school and teacher education: A focus on a societal perspective

Takashi Kawakami

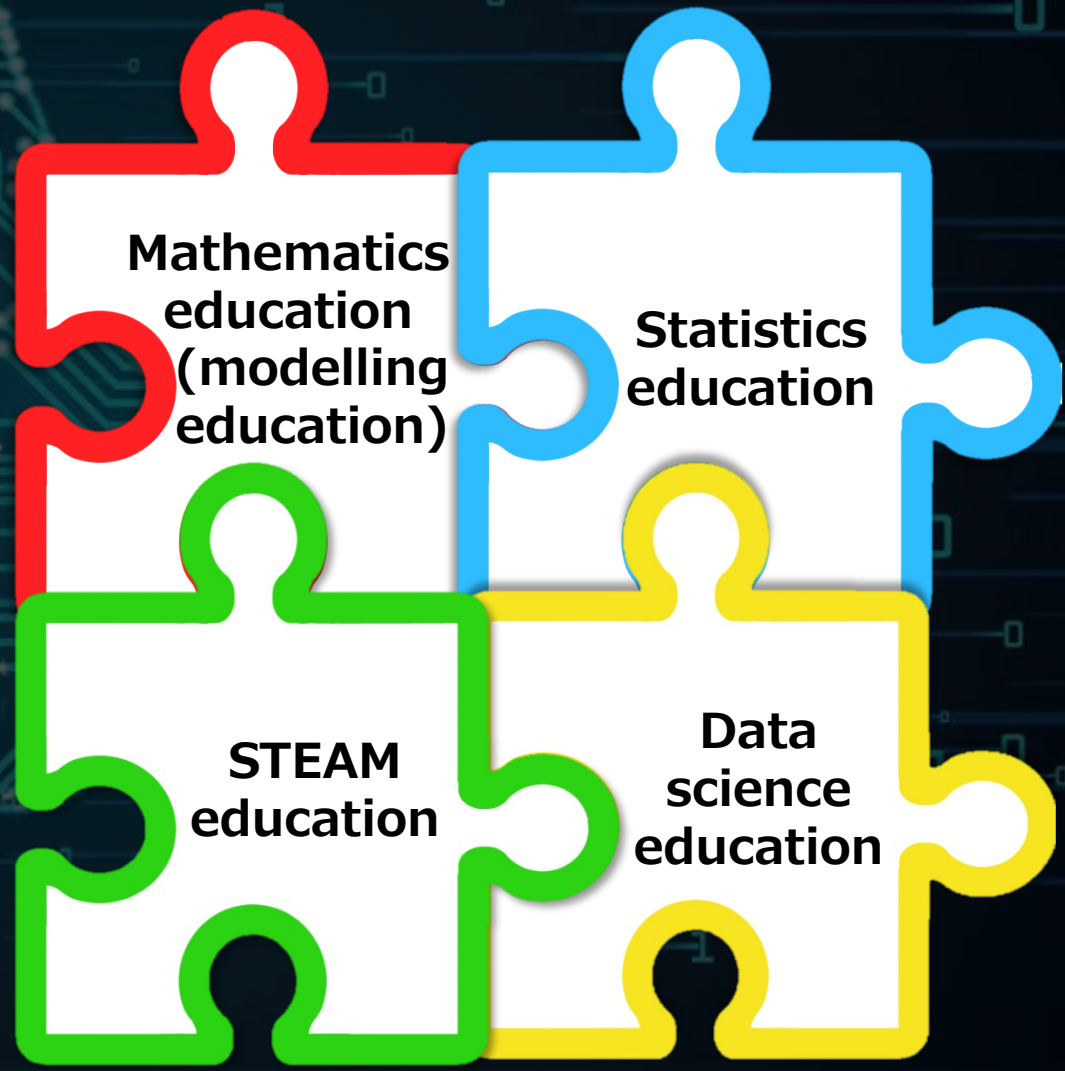
Utsunomiya University, Japan

ProDaBi
Projekt Data Science und Big Data in der Schule

宇都宮大学
UTSUNOMIYA UNIVERSITY

My context

Graduated from Hyogo University for Teacher Education (Japan) in March 2023 with a doctoral thesis on *Data-Driven Modelling in School Mathematics* (Supervised by Prof. Dr. Akihiko Saeki)



Targeting school education

Targeting teacher education

Outline

Data-driven modelling (DDM) approach with deterministic (mathematical) and stochastic (statistical) models at its core in school and teacher education: A focus on a societal perspective

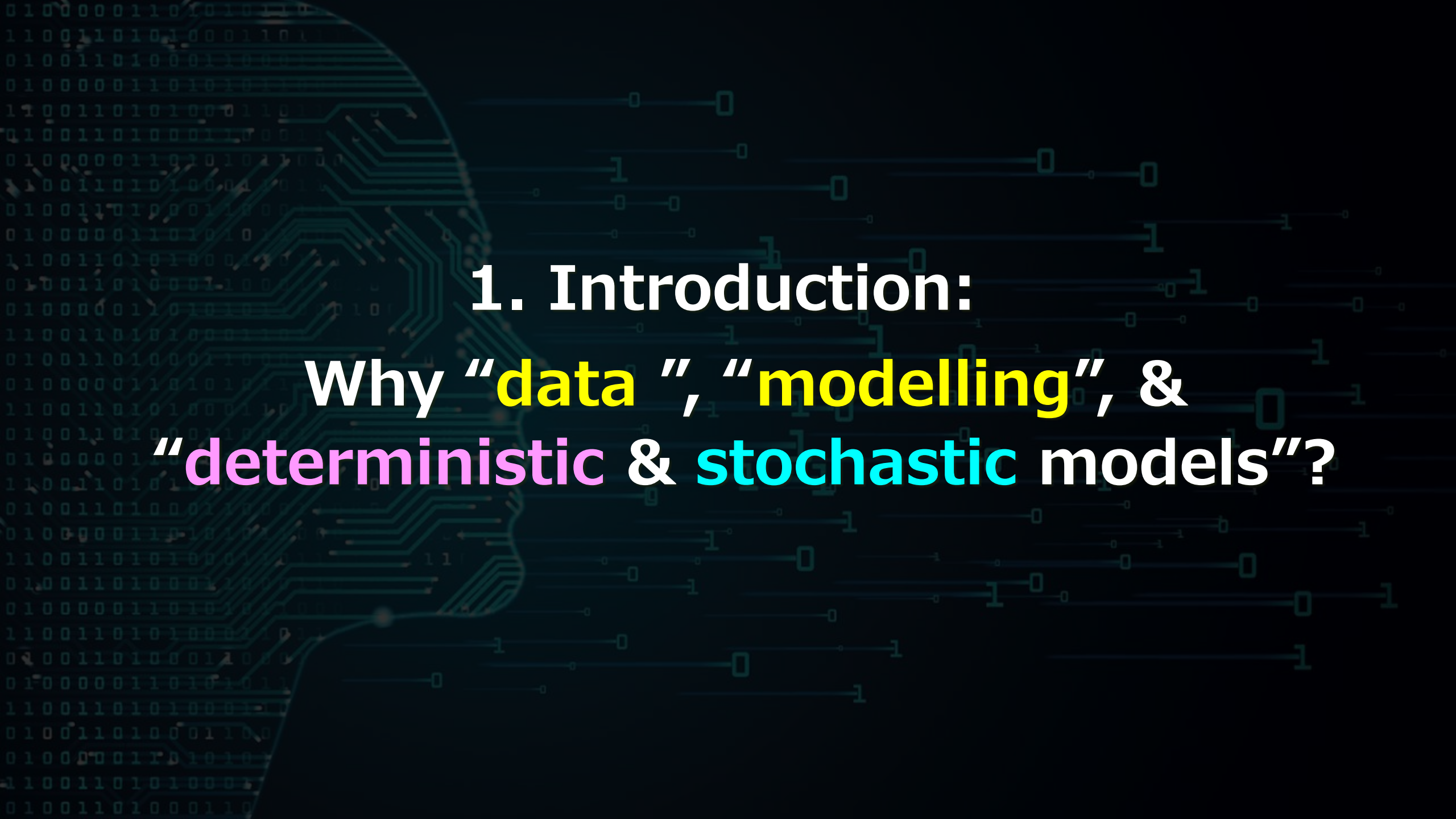
1. Introduction:
Why "data ", "modelling", & "deterministic & stochastic models"?

2. Overview of common discourses between mathematical modelling education and statistics/data science education research

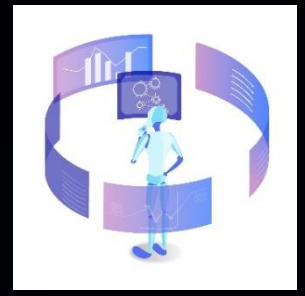
3. Data-driven modelling (DDM) framework for primary & secondary schools

4. A case of societal DDM practice with pre-service primary & secondary school teachers

5. Summary

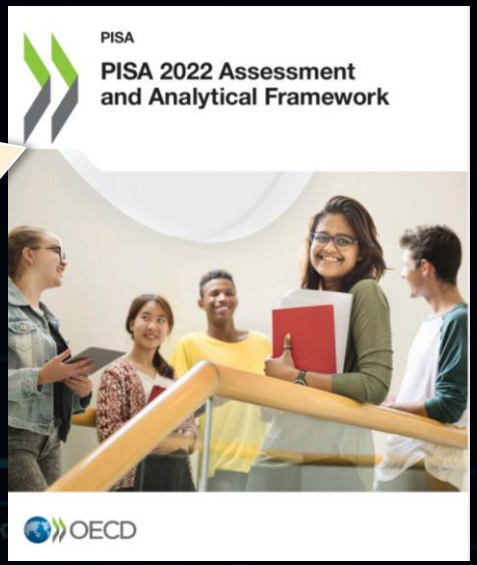

1. Introduction:
Why “data ”, “modelling”, &
“deterministic & stochastic models”?

Why "data ", "modelling", & "deterministic & stochastic models"?



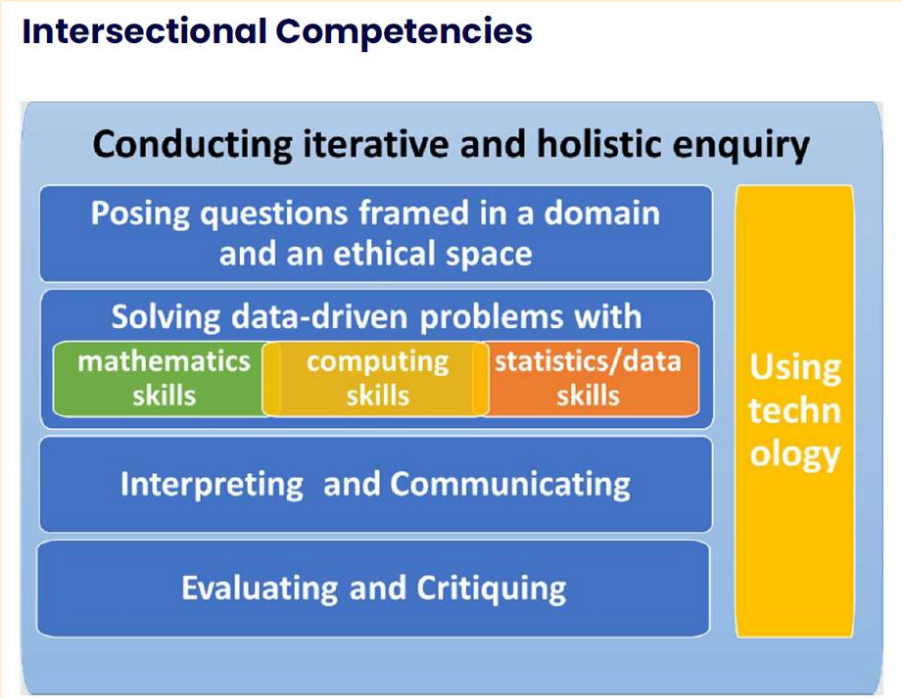
- The role of data in today's data-driven and AI society is increasing
- Data is often central to the creation and validation of models (representations) with mathematics and statistics (e.g., Biehler, 2022; Gal, 2024; Phannkuch et al., 2018; Ridgway et al., 2022; Siller et al., 2024)
- Future citizens are required for a comprehensive competency to flexibly use and think critically about both deterministic and non-deterministic/stochastic models in their daily life such as media and AI (e.g., Engel, 2024; Gal & Geiger, 2022; Geiger et al., 2023; OECD, 2023; Smith et al., 2023)

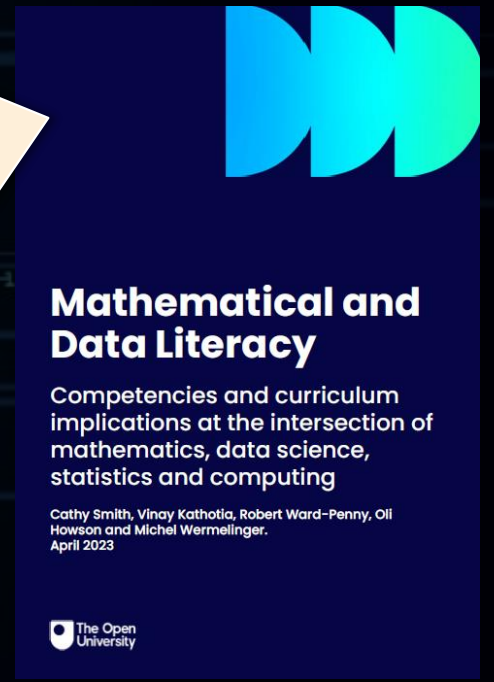
"Two aspects of mathematical reasoning are especially important in today's world and in defining the PISA items. One is deduction from clear assumptions (deductive reasoning), which is a characteristic feature of mathematical process. …The second important dimension is statistical and probabilistic (inductive) reasoning." (OECD, 2023, p. 28)



"Within this framework, problem solving can draw on 'toolkits' in mathematics, computing and statistics/data science…"

(Smith et al., 2023, p. 7)





Why “data ”, “modelling”, & “deterministic & stochastic models”?

- In national-level school mathematics curricula (e.g., Australia, Germany, Japan, USA), both mathematical modelling and data and statistical investigation have been emphasized.
- In research, recent discussions on the boundary between mathematical modelling education and statistics/data science education have focused on modelling with statistics and mathematics at its core (e.g., Ärlebäck & Kawakami, 2023; Gal & Geiger, 2022; Kawakami & Ärlebäck, 2024; Kazak, Fujita, & Turmo, 2023; Leavy et al., 2018; Lehrer & English, 2018; Makar, Fry, & English, 2024).

2. Overview of common discourses between mathematical modelling education and statistics/data science education research

The slides in this section are based on the slides I presented as part of the ICME-15 survey team

(Biehler et al., 2024)

15th International Congress on Mathematical Education

7-14 July 2024 • ICC Sydney, Australia

Come and be counted

Survey 3: Statistics and data science education as a vehicle for empowering citizens

IPC Liaison: Katie Makar

Leader: Prof. Rolf Biehler, Universitat Paderborn (Germany)

Member: Prof. Takashi Kawakami, Utsunomiya University (Japan)

Member: Dr. Erna Lampen, Stellenbosch University (South Africa)

Member: Dr. Lucia Zapata-Cardona, Universidad De Antioquia (Columbia)

Member: A/Prof Travis Weiland, University of Houston (United States)

The team will also report in an IASE webinar in May 2025.

Three common discourses between math modelling & statistics/data science educ. research from 2020

Discourse 1

Modelling that integrates statistics & mathematics

Discourse 2

Interdisciplinary modelling

Discourse 3

Societal modelling

A long version of this review will be available in 2025.

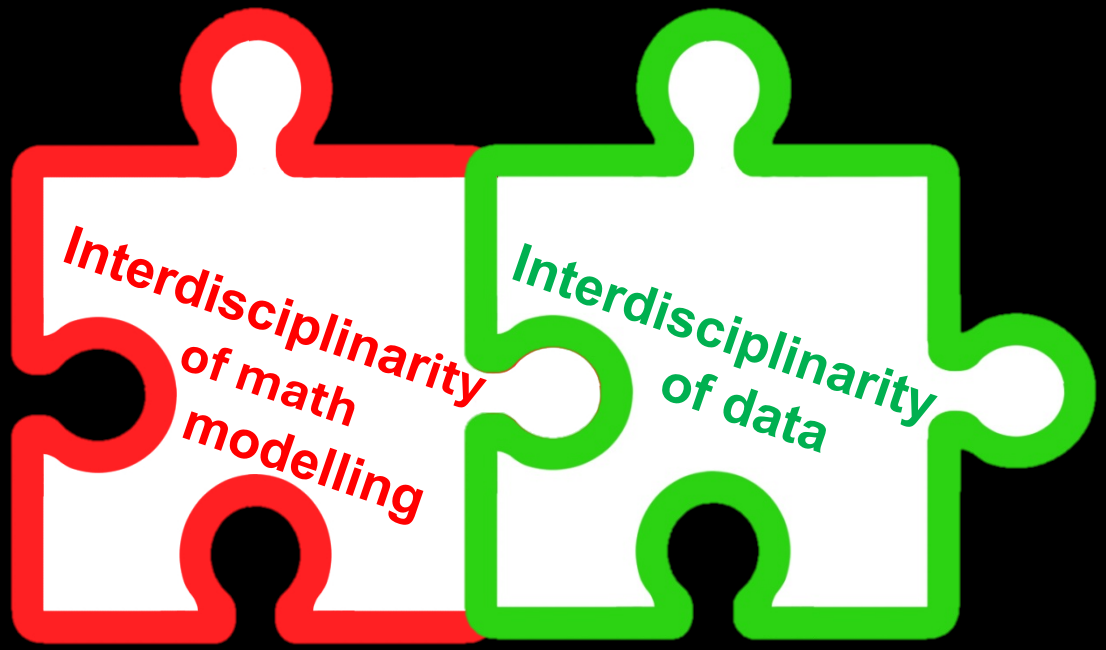
Role of data highlighted by empirical studies in Discourse 1

- Elicit contextual knowledge, story, and patterns (trends and variation) behind or in the data to build models to make better predictions and decisions (e.g., Kazak et al., 2023; McLean et al., 2023; Stillman & Brown, 2023; Van Dijke-Droogers et al., 2021)

- Generate the hypothetical model for prediction with one data set and then test it on another one (e.g., Ärlebäck, Frejd, & Doerr, 2021; Dvir & Ben-Zvi, 2023; Kawakami & Mineno, 2021)

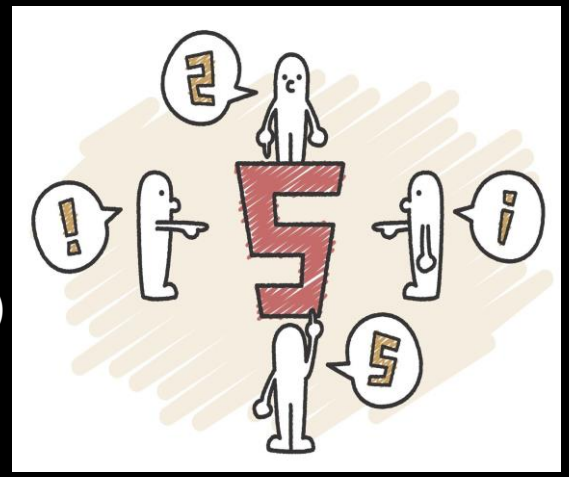
Enhancing intardicplinarity in Discorse 2

- Interdisciplinary data-rich modelling with not only statistics and mathematics, but also other disciplines/subjects to promote also “STEM literacy” (e.g., Bybee, 2018)
  - The inherent interdisciplinarity of mathematical modelling (e.g., English, 2016; Maass et al., 2019; Pollak, 1977; Stillman et al., 2023)
  - The inherent interdisciplinarity of variation and ethical aspects in data (e.g., Lehrer & Schauble, 2002; Makar, Fry, & English, 2023; Watson, Fitzallen, & Chick, 2020)



Developing multifaceted views and knowledge in Discourse 2

- The back-and-forth movement between deterministic views, stochastic views, and other STEM views such as scientific views, design/engineering views (e.g., Aridor, Dvir, Tsybulsky, & Ben-Zvi, 2023)
- Engagement of non-standard data, data privacy, and ethical issues in age-appropriate ways (e.g., Makar, Fry, & English, 2024)
- Understanding of the role of uncertainty in generating data-based interdisciplinary knowledge (e.g., Lehrer, Wisittanawat, & Schauble, 2024)



Developing citizenship in Discorse 3

3 Societal modelling

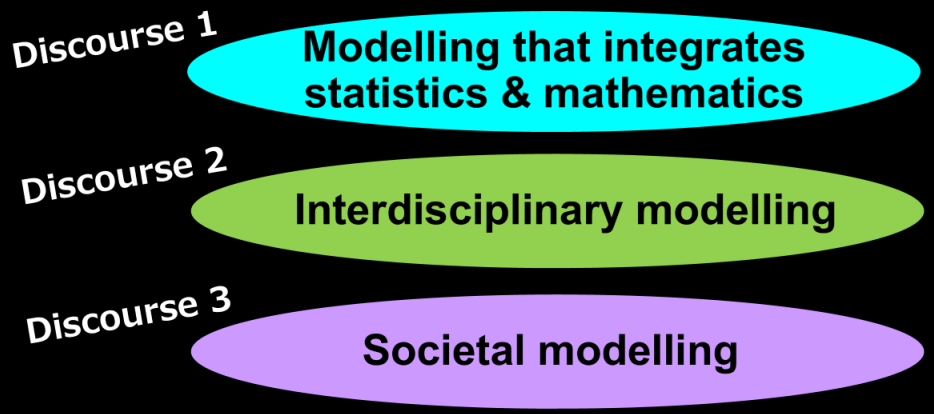
- **Societal data-rich modelling** with statistics and mathematics to promote citizenship and critical thinking
- The social, critical, and prescriptive/performative aspects of mathematical modelling (e.g., Barbosa, 2006; Davis & Hersh, 1986; Niss, 2015; Skovsmose, 2024)
  - **Prescriptive modeling** (Niss, 2015): "*pave the way for taking action based on decisions resulting from a certain kind of mathematical considerations, in other words to change the world*" (p. 69)
- Use of global, social, political, ethical, and daily life contexts to create authentic data (science) practices

Context example in Discourse 3

3 Societal modelling

- *COVID-19 or epidemics* (e.g., Maass et al., 2023)
- *Climate change* (e.g., Kazak et al., 2023; Steffensen & Kacerja, 2021; Zapata-Cardona & Martínez-Castro, 2023)
- *Pandemic-related media items* (Gal & Geiger, 2022)
- *Reliability of public data sets* (e.g., Wilkerson et al., 2022)
- *Mapping crime in the regions* (Andersson & Register, 2023)
- *Social justice in fair distribution of school funding* (Jung & Wickstrom, 2023)
- *Trash production* (Rosa, Orey, & de Sousa Mesquita, 2023)

Three common discourses between math modelling & statistics/data science educ. research from 2020

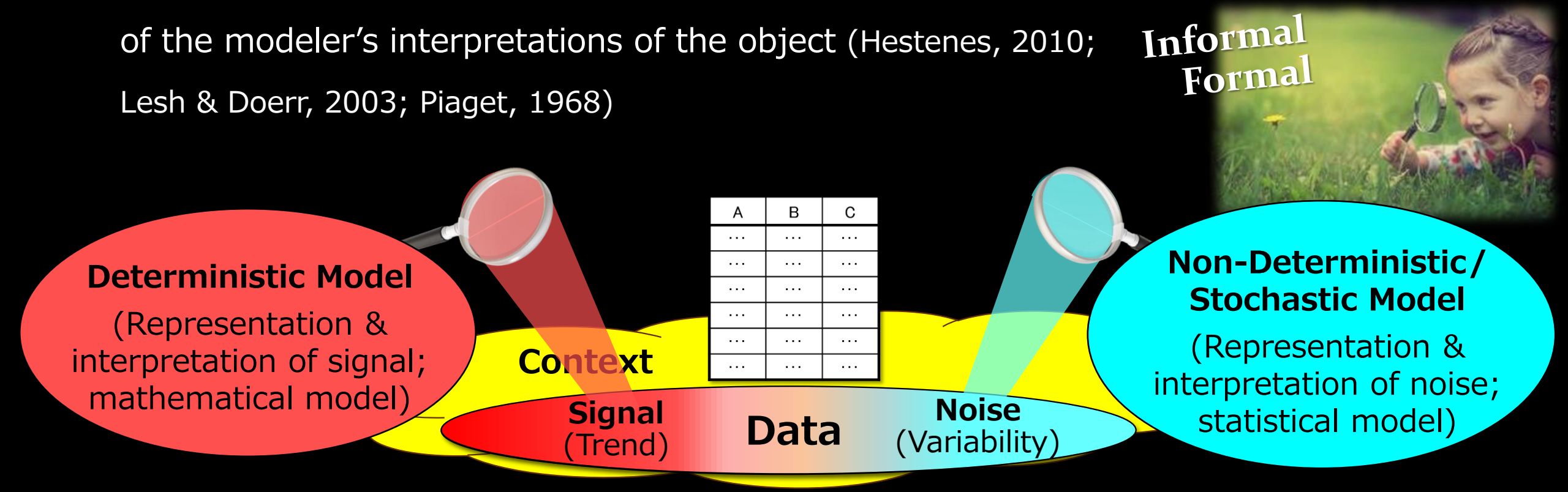


- In both mathematical modelling education research and statistics/data science education research, “data and modelling“ is being developed as a common but distinct research topic. This leads to a theoretical separation of modelling in both research areas.
- Very few comprehensive theoretical and practical frameworks to share *“the research concerns of mathematics and statics[/data science] educators with respect to modelling”* (Makar & Rubin, 2018, p. 289) .

3. Data-driven modelling (DDM) framework for primary & secondary schools

“Data” & “model” in the framework

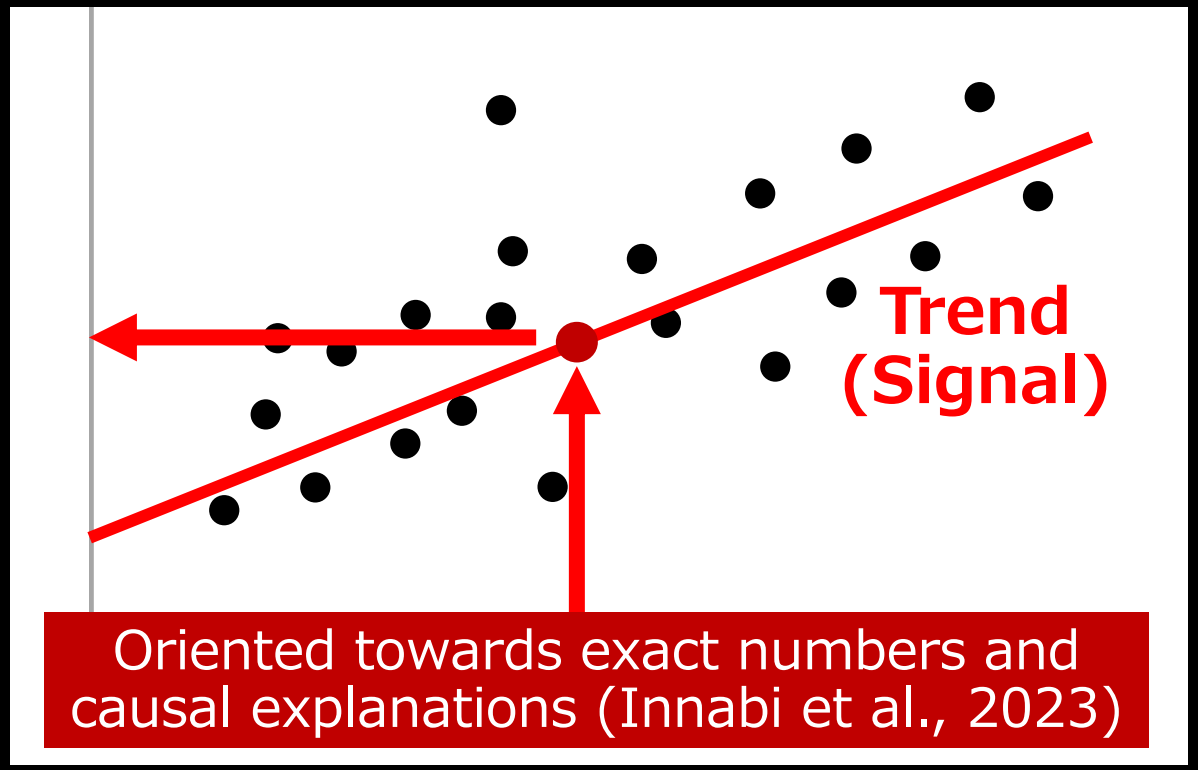
- **Data**: Closely related to real-world context and consist of signal (trend) and noise (variability) (Cobb & Moore, 1997; Dvir & Ben-Zvi, 2023; Konold & Pollatsek, 2002)
- **Model**: A representation of the essential characteristics of an object and a reflection of the modeler’s interpretations of the object (Hestenes, 2010; Lesh & Doerr, 2003; Piaget, 1968)



Deterministic Model

(Representation & interpretation of signal; mathematical model)

(ex.) Regression line for linear function

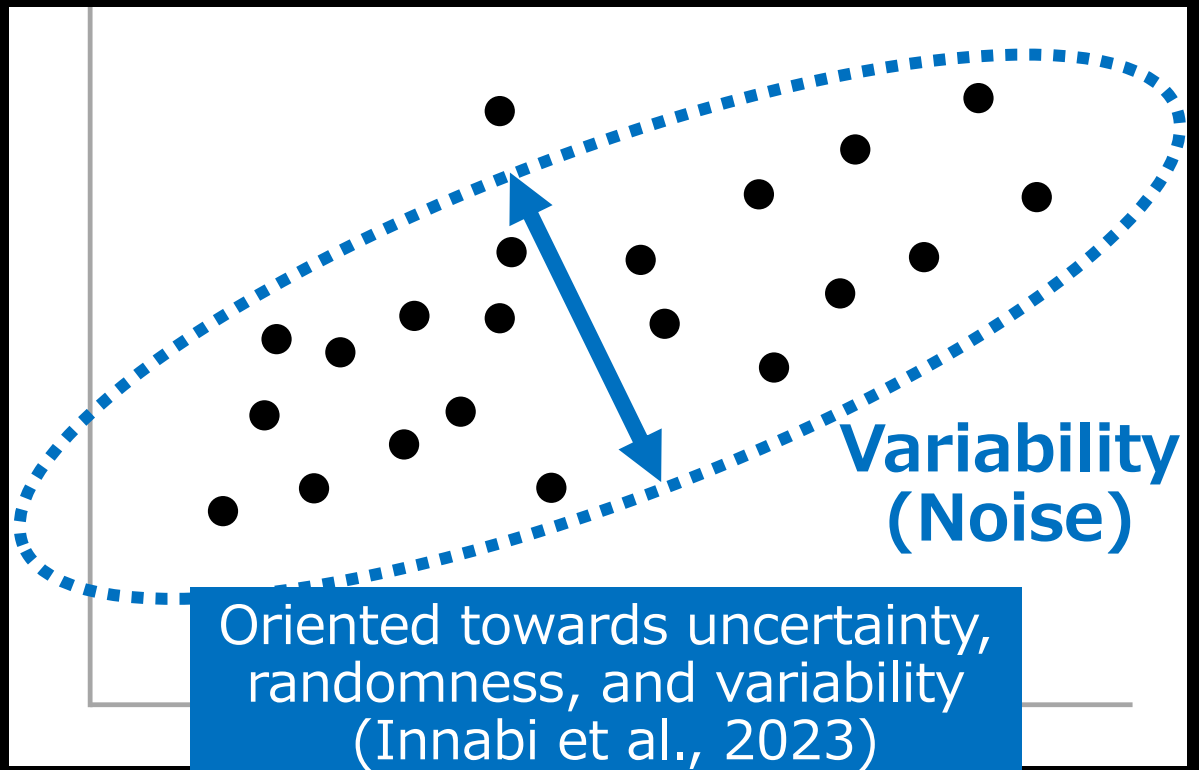


Oriented towards exact numbers and causal explanations (Innabi et al., 2023)

Non-Deterministic/ Stochastic Model

(Representation & interpretation of noise; statistical model)

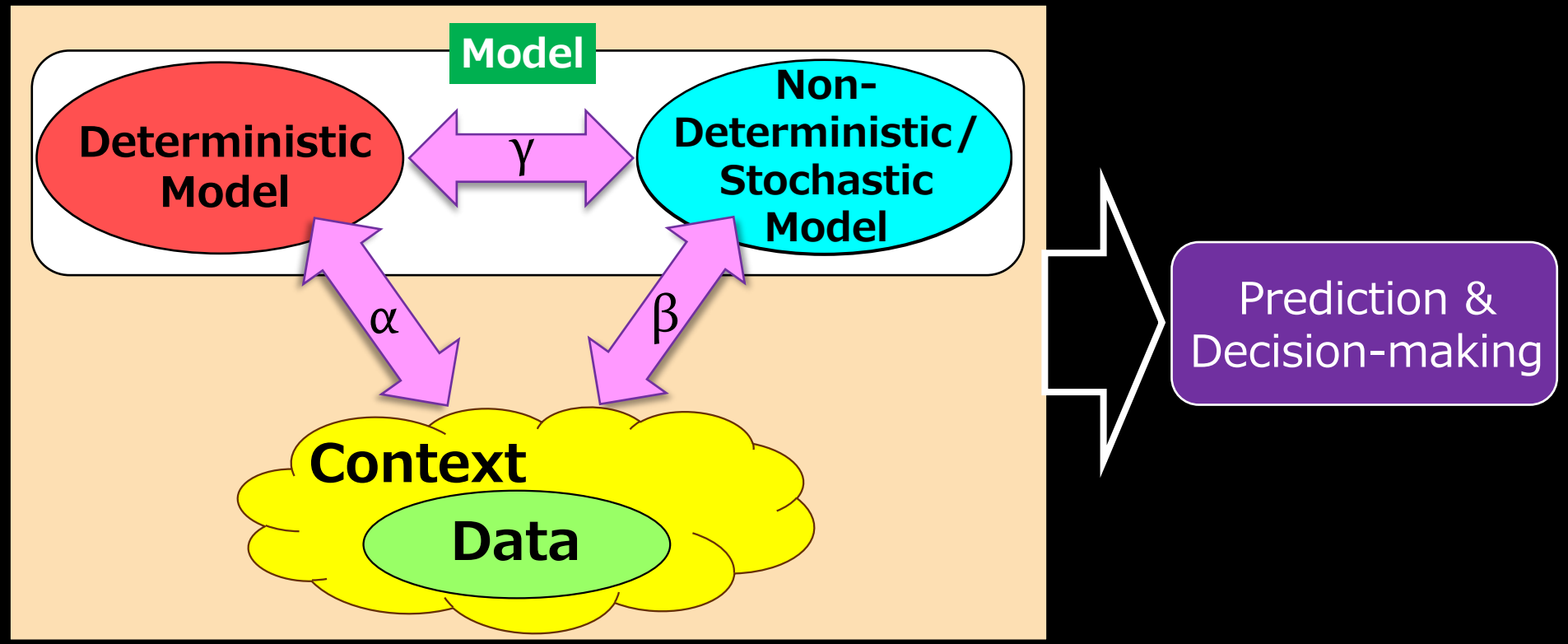
(ex.) Representation of the width regression



Oriented towards uncertainty, randomness, and variability (Innabi et al., 2023)

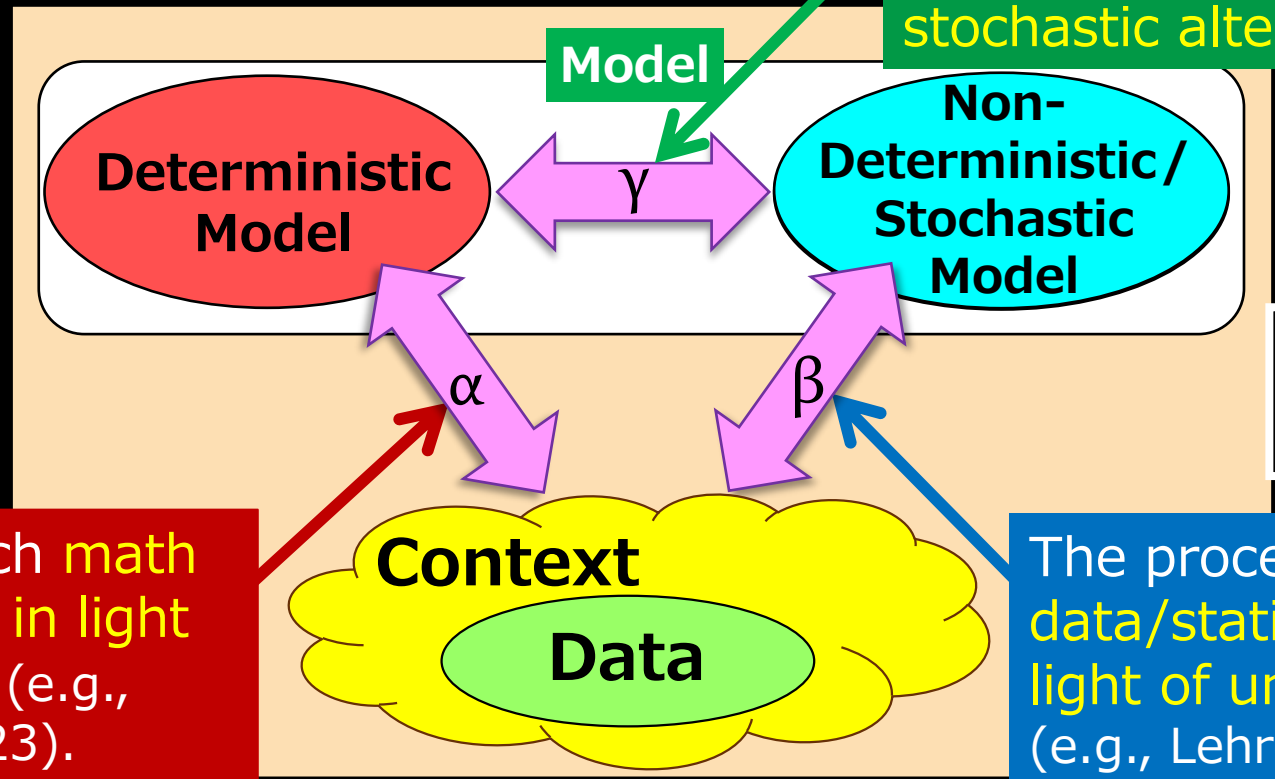
Data-driven modelling (DDM) framework for primary & secondary schools (Kawakami, 2023; Kawakami & Saeki, 2022)

- Activities involving the generation, validation, and modification of deterministic (mathematical) and/or non-deterministic/stochastic (statistical) models, based on data to make better predictions and decision-making



Data-driven modelling (DDM) framework for primary & secondary schools (Kawakami, 2023; Kawakami & Saeki, 2022)

The process of evaluation of an initial deterministic model could lead to generation of a non-deterministic/ stochastic alternative and vice versa.

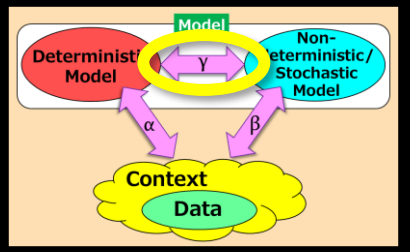


Prediction & Decision-making

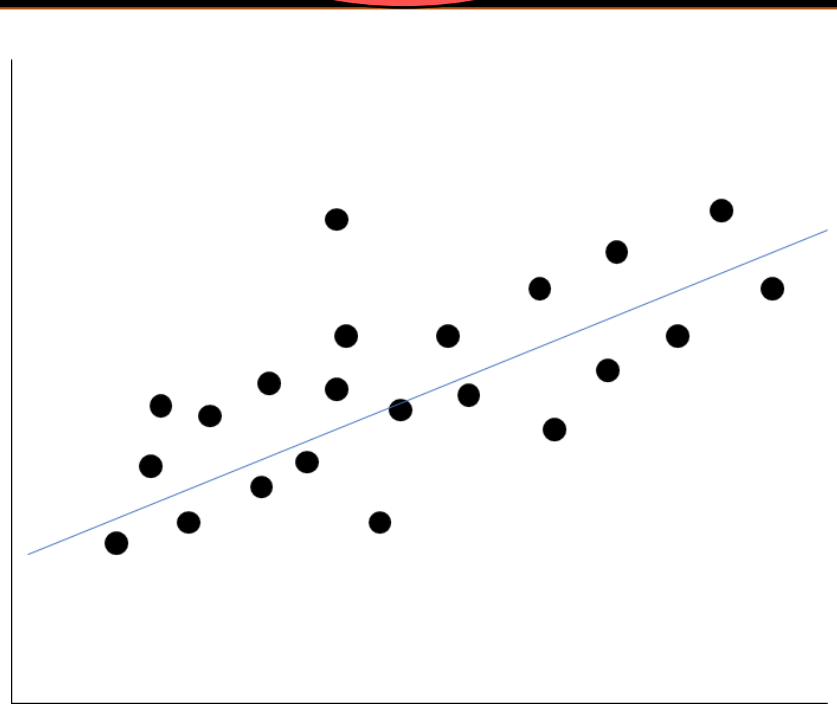
The process on which math modelling with data in light of certainty focuses (e.g., Stillman & Brown, 2023).

The process on which data/statistical modelling in light of uncertainty focuses (e.g., Lehrer & English, 2018).

Typical example of transition γ



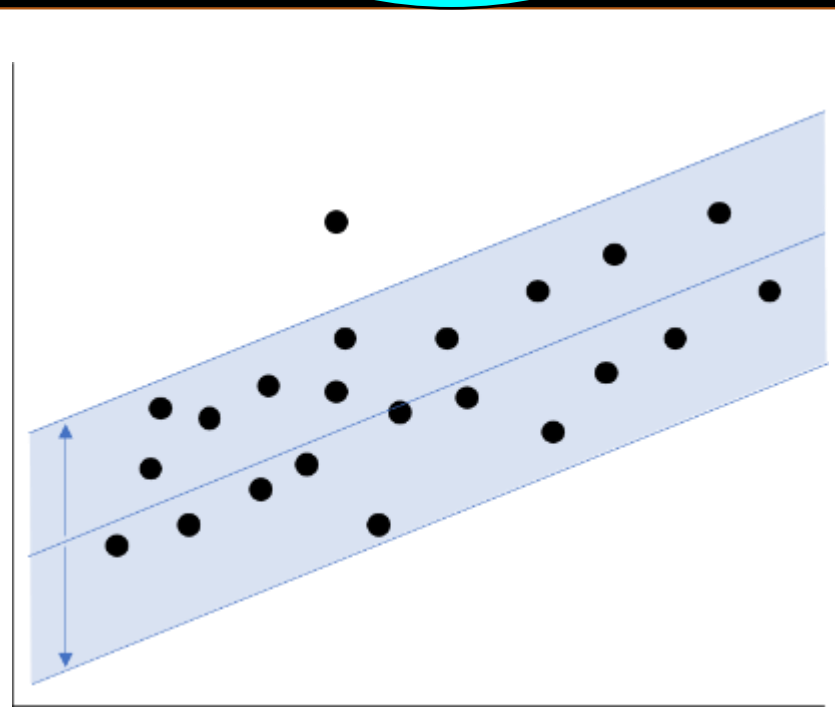
Deterministic Model



Through non-deterministic/stochastic representation and interpretation

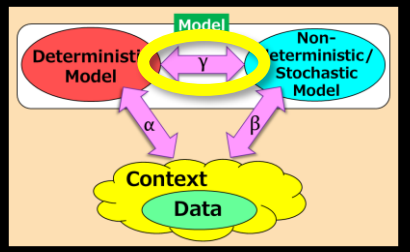
Through deterministic representation and interpretation

Non-Deterministic/ Stochastic Model



The case of transition γ in Grade 2 classroom

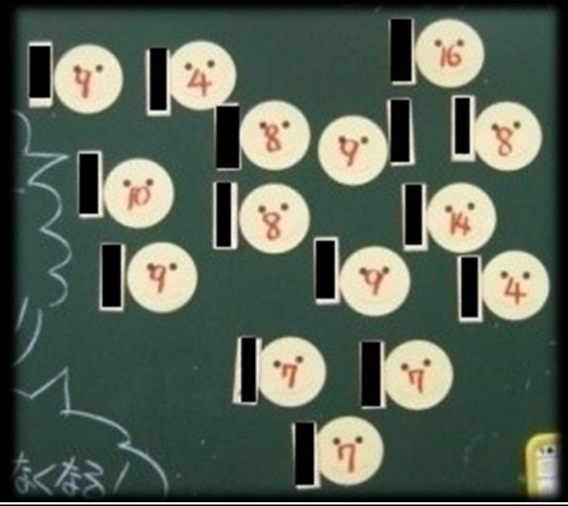
(Kawakami, 2022)



- **Participants:** Grade 2 students (aged 7-8) in Japan
- **Context:** Baby teeth loss prediction (cf. Ben-Zvi & Sharett-Amir, 2005)

- **Task (partially):** We plan to predict the distribution of baby teeth loss for the whole class and for children in other classes. We have collected data on the number of baby teeth that have fallen out for the 15 friends in the class.

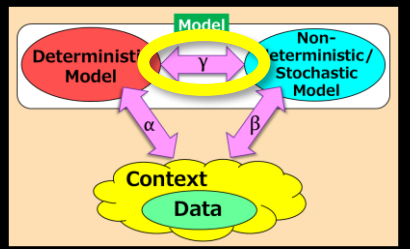
How should the face icons be arranged to make it easier to see how many baby teeth have fallen out for the 15 friends?

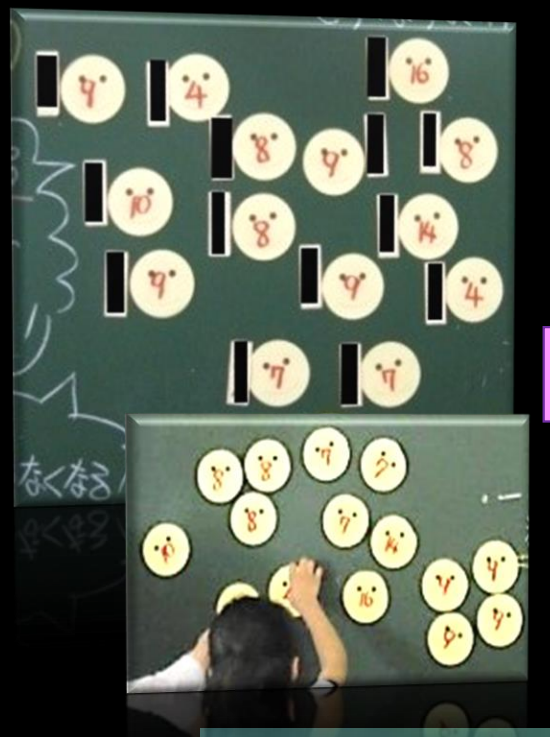


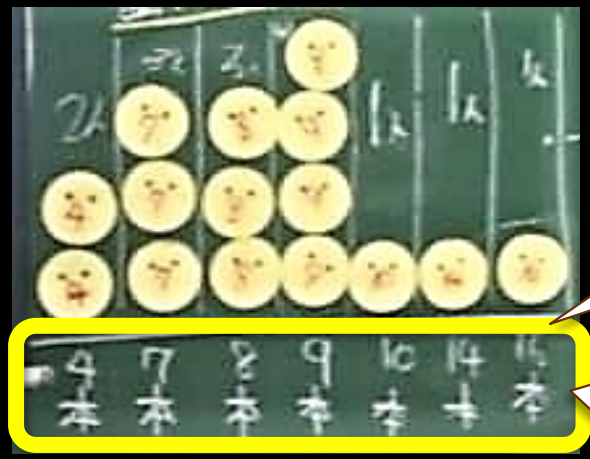
Kawakami, T. (2022). The role of models in promoting informal statistical inferences of lower grade children: Focusing on data modelling processes. *Journal of Science Education in Japan*, *46*(2), 125–140.

The case of transition γ in Grade 2 classroom

(Kawakami, 2022)

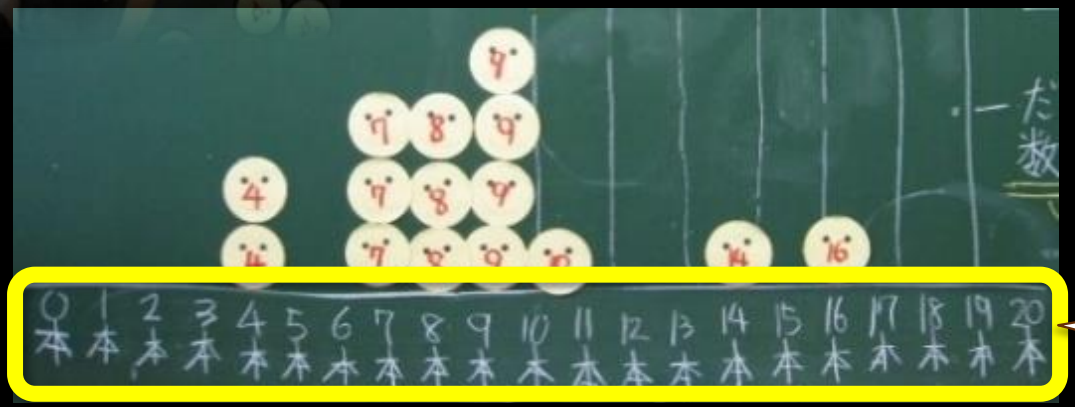






"What values should we write on the horizontal axis?"

"There may be friends in the other class or Grade 1 students who have not yet lost their baby teeth (a value of 0)."

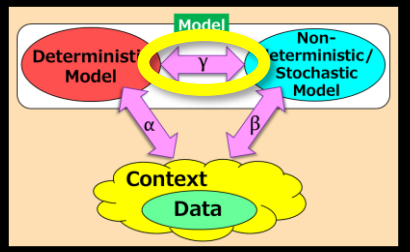


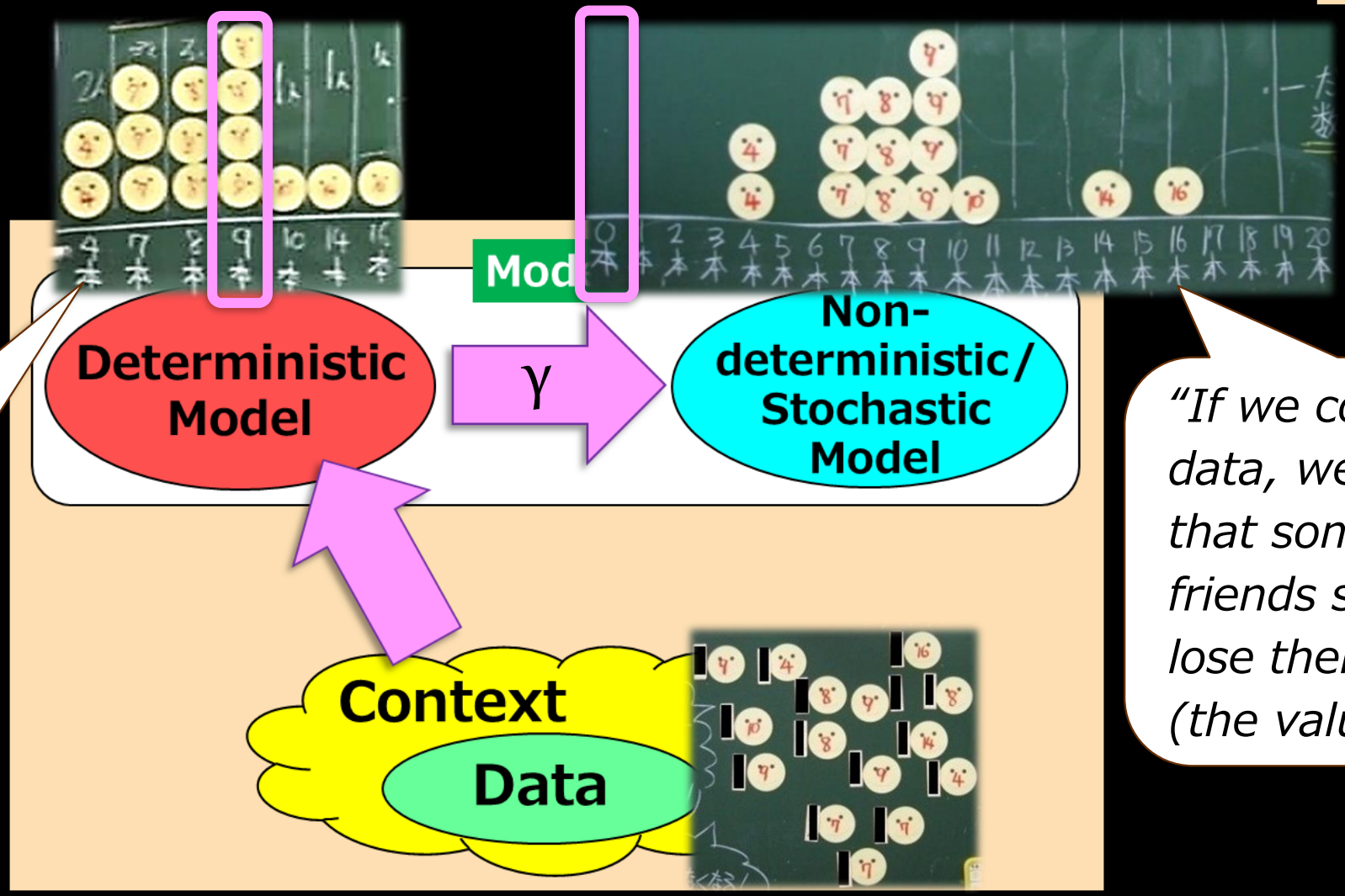
Stochastic interpretation

The interpretation of the horizontal axis changes to 'possible' values.

The case of transition γ in Grade 2 classroom

(Kawakami, 2022)

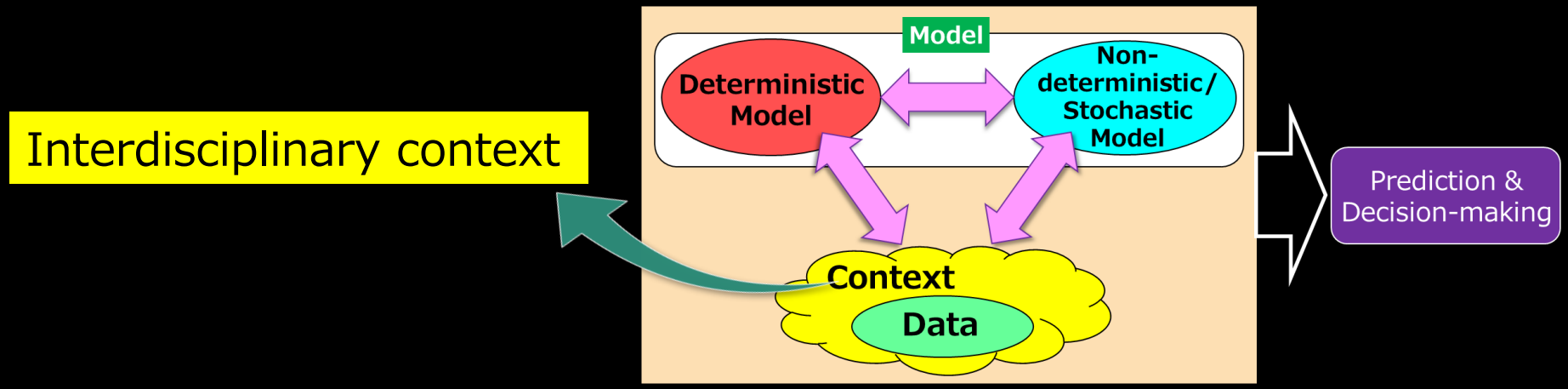




"The number of people with nine missing baby teeth is the highest."

"If we collect more data, we might find that some of our friends still do not lose their baby teeth (the value is 0)."

Three types of DDM approaches

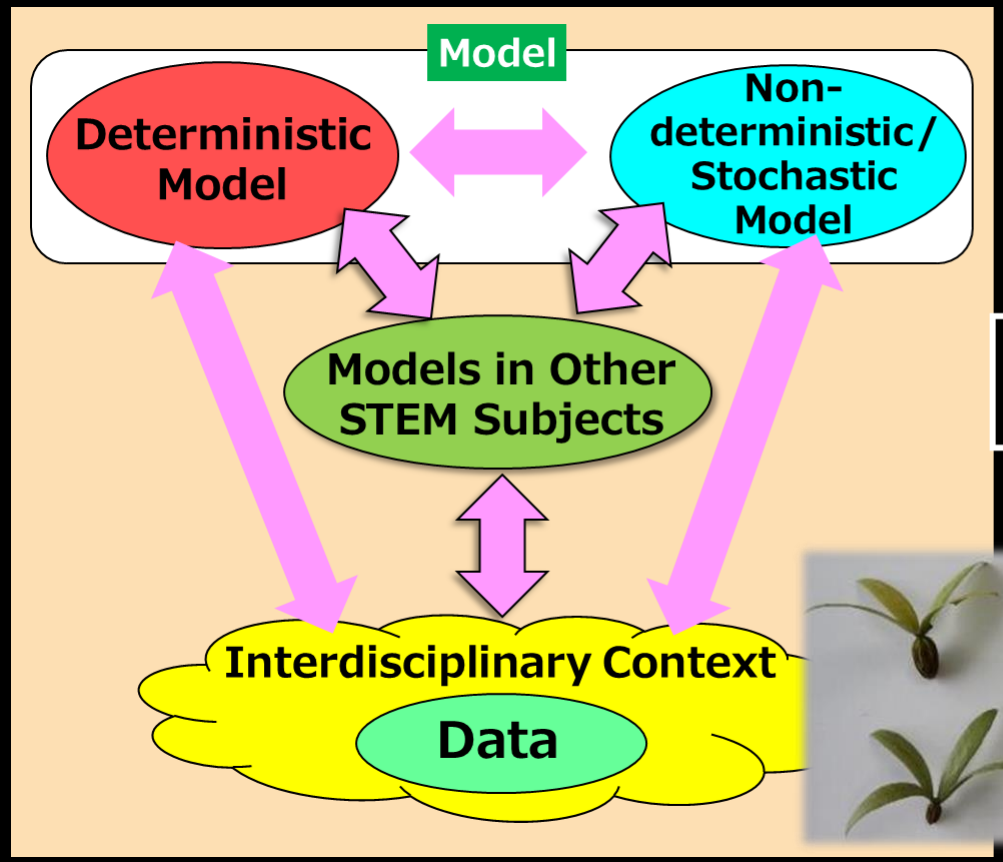


1. DDM with mathematics and statistics at its core from mathematics and statistics education perspectives (Kawakami, 2022, 2023; Kawakami & Saeki, 2022)

2. DDM with mathematics and statistics, along with other disciplines and subjects from an interdisciplinary perspective (Kawakami & Nishimura, 2024; Kawakami & Saeki, 2024a)

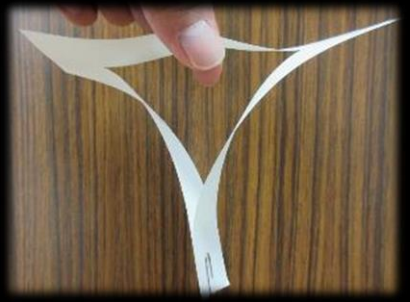
Interdisciplinary DDM with Grade 4 (Kawakami & Saeki, 2024a)

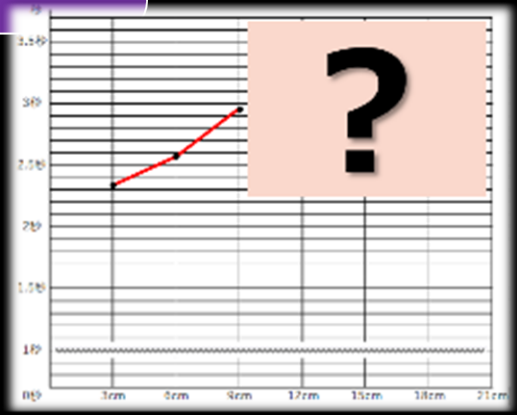
- **Context:** Redesign of the shape of the seed to maximise flight time (cf. Watson et al., 2020)



Predictions along with both scientific and design models

Prediction & Decision-Making





Kawakami, T., & Saeki, A. (2024a). Extending data-driven modelling from school mathematics to school STEM education. In J. Anderson, & K. Makar (Eds.), *The contribution of mathematics to school STEM education: Current understandings* (pp. 221-239). Springer.

Three types of DDM approaches

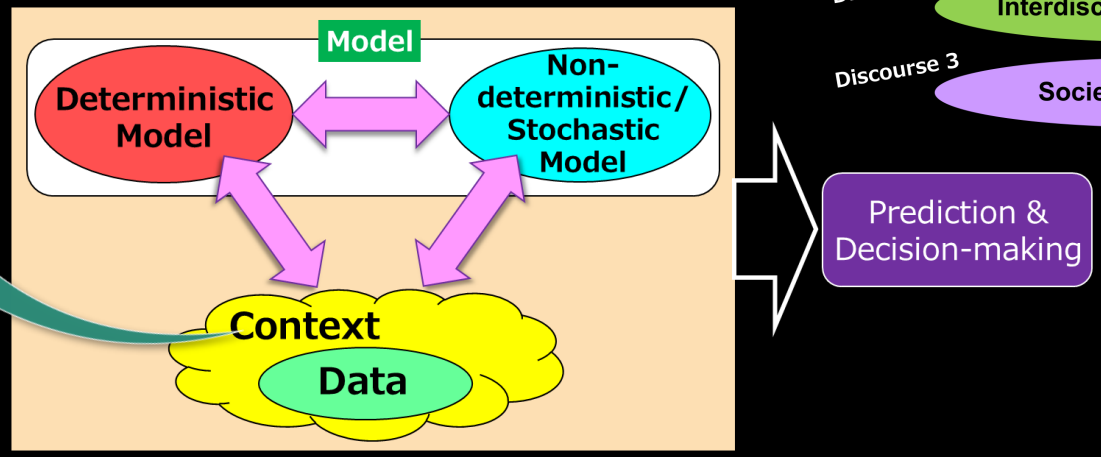
Three common discourses between math modelling & statistics/data science educ. research from 2020

Discourse 1: Modelling that integrates statistics & mathematics

Discourse 2: Interdisciplinary modelling

Discourse 3: Societal modelling

Civic, social, or societal context

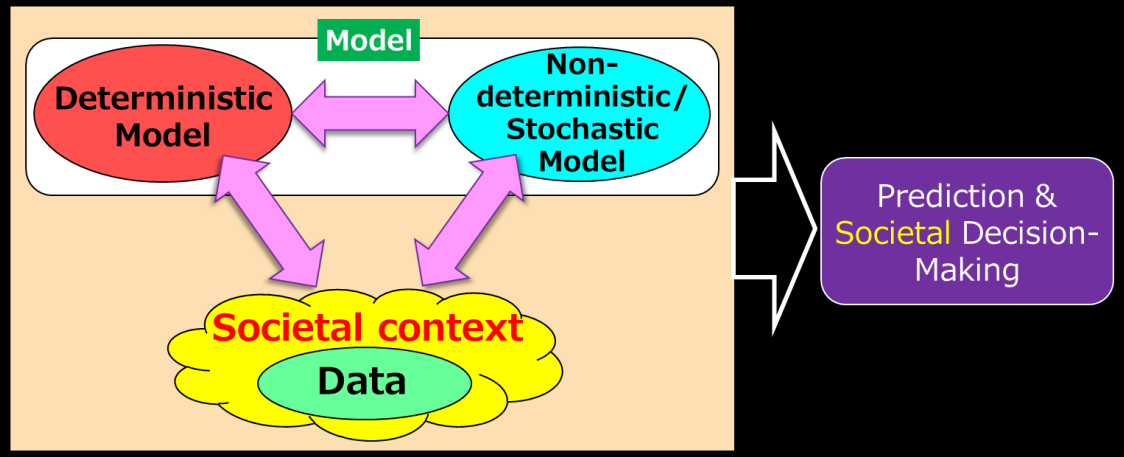


1. DDM with mathematics and statistics at its core from mathematics and statistics education perspectives (Kawakami, 2022, 2023; Kawakami & Saeki, 2022)

2. DDM with mathematics and statistics, along with other disciplines and subjects from an interdisciplinary perspective (Kawakami & Nishimura, 2024; Kawakami & Saeki, 2024a)

3. DDM with a focus on social decision-making using mathematics and statistics from a societal perspective (Kawakami & Saeki, 2024b)

Three types of DDM approaches



1. **DDM with mathematics and statistics at its core from mathematics and statistics education perspectives** (Kawakami, 2022, 2023; Kawakami & Saeki, 2022)

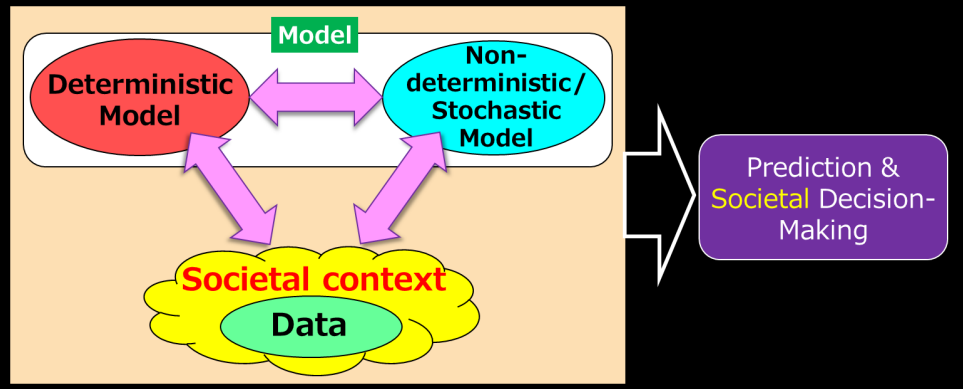
2. **DDM with mathematics and statistics, along with other disciplines and subjects from an interdisciplinary perspective** (Kawakami & Nishimura, 2024; Kawakami & Saeki, 2024a)

3. **DDM with a focus on social decision-making using mathematics and statistics from a societal perspective** (Kawakami & Saeki, 2024b)

4. A case of societal DDM practice with pre-service primary & secondary school teachers

Setting of societal DDM practice (Kawakami & Saeki, 2024b)

- **Participants:** 29 pre-service primary and secondary mathematics teachers (aged 20－21 years) in Japan
- A teaching experiment of 8 online classes and 3 report assignments (Oct. 2020－Feb. 2021, during the pandemic of COVID-19)
  - ➢ To experience basic societal DDM and become aware of the societal benefits and risks of reconstructing reality in society through societal DDM
- **Context:** Data-informed decision-making on the COVID-19 pandemic in Japan



Kawakami, T., & Saeki, A. (2024b). Roles of mathematical and statistical models in data-driven modelling: A prescriptive modelling perspective. In H.S. Siller, V. Geiger, & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 595-605). Springer.

COVID-19 task (Partially)

in the final report assignment

47 prefectures in Japan

Assume you are in the position of the government. In which prefecture would you declare a state of emergency? Decide based on actual data (as of January 12, 2021) and discuss the process that led to that decision in a Word file. The actual data are stored in data analytic education software CODAP. You may use Excel or other programs as needed.

(Kawakami & Saeki, 2024b, p. 599)

CODAP (http://codap.concord.org)

- **Data** such as *population*, *number of COVID-19 cases and admissions*, *number of beds for COVID-19 patients and critical cases* in Japan's 47 prefectures

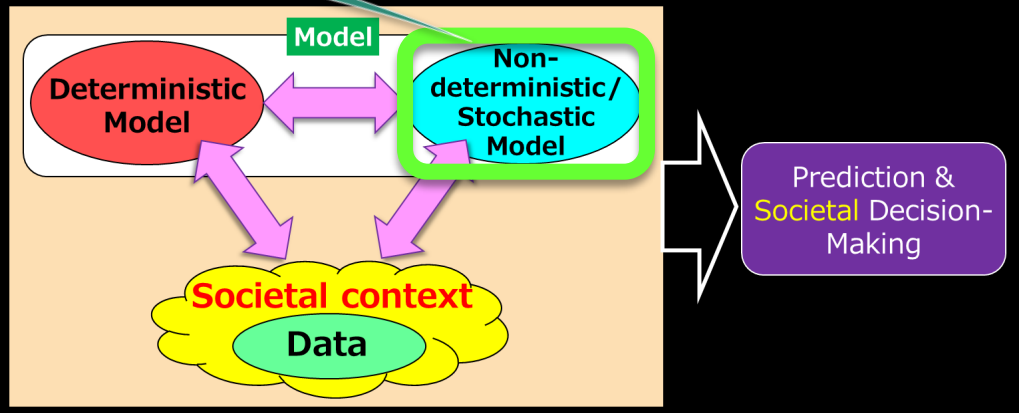
都道府県のCOVID-19データ（1月12日現在）

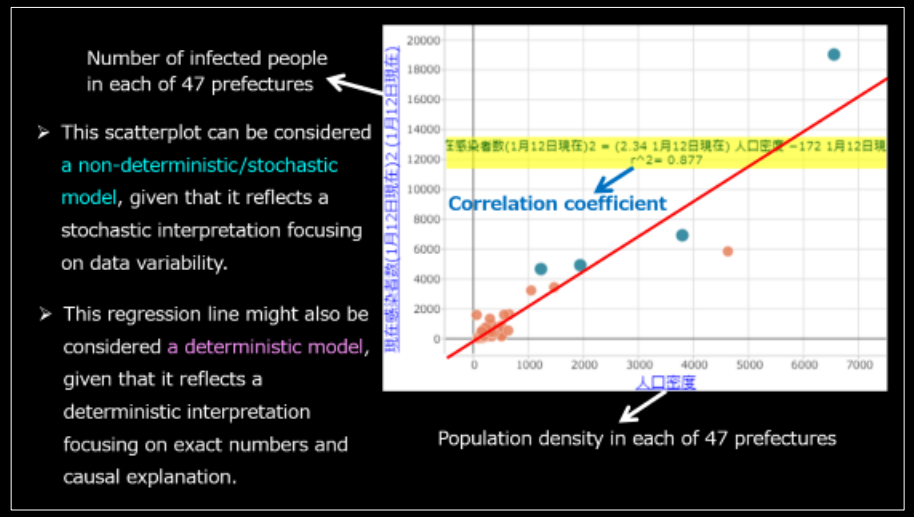
都道府県データ (47データ)

索引	No	県	境界	海	人口	面積 (km²)	現在感染者数(3月31日現在)2	現在感染者数(4…4月30日現在)	現在感染者数(5…(5月31日現在)2	現在感染者数(7…月31日現在)2	現在感染者数…(9月30現在)	現在感染者数…(11/29現在)	現在感染者…(1/12現在)	感染者数感染…(3月31日現
1	1	北海道		有り	5286000	83424	38	485	195	82	138	2323	1588	
1	2	青森県		有り	1263000	9646	8	11	1	2	1	16	70	
2	3	岩手県		有り	1241000	15275	0	0	0	4	0	95	65	
3	4	宮城県		有り	2316000	6859	6	36	0	19	45	167	497	
4	5	秋田県		有り	981000	11638	4	5	0	2	0	17	28	
5	6	山形県		有り	1090000	6652	1	20	3	1	2	31	60	
6	7	福島県		有り	1864000	13784	4	49	6	6	37	54	355	
1	8	茨城県		有り	2877000	6097	24	119	7	68	37	406	749	
2	9	栃木県		無し	1946000	6408	12	33	14	56	49	107	1335	
3	10	群馬県		無し	1952000	6362	18	100	12	30	55	200	589	
4	11	埼玉県		無し	7330000	3768	75	572	66	504	264	1386	4925	
5	12	千葉県		有り	6255000	5083	143	632	55	370	272	966	4667	
6	13	東京都		有り	13822000	2109	473	2730	416	2742	2177	4407	19029	
7	14	神奈川県		有り	9177000	2416	111	474	222	342	570	1928	6923	
1	15	新潟県		有り	2246000	10364	22	36	5	22	8	103	159	
2	16	富山県		有り	1050000	2046	2	157	16	6	10	25	176	
3	17	石川県		有り	1143000	4186	9	185	54	19	43	17	164	
4	18	福井県		有り	774000	4191	20	44	4	14	4	22	40	
5	19	山梨県		無し	817000	4254	5	16	4	17	12	48	175	
6	20	長野県		無し	2063000	13104	5	48	7	23	8	135	530	
7	21	岐阜県		無し	1997000	9769	24	66	3	123	31	184	703	

Pre-service teachers (PTs)' use of models given in the DDM framework (Kawakami & Saeki, 2024b)

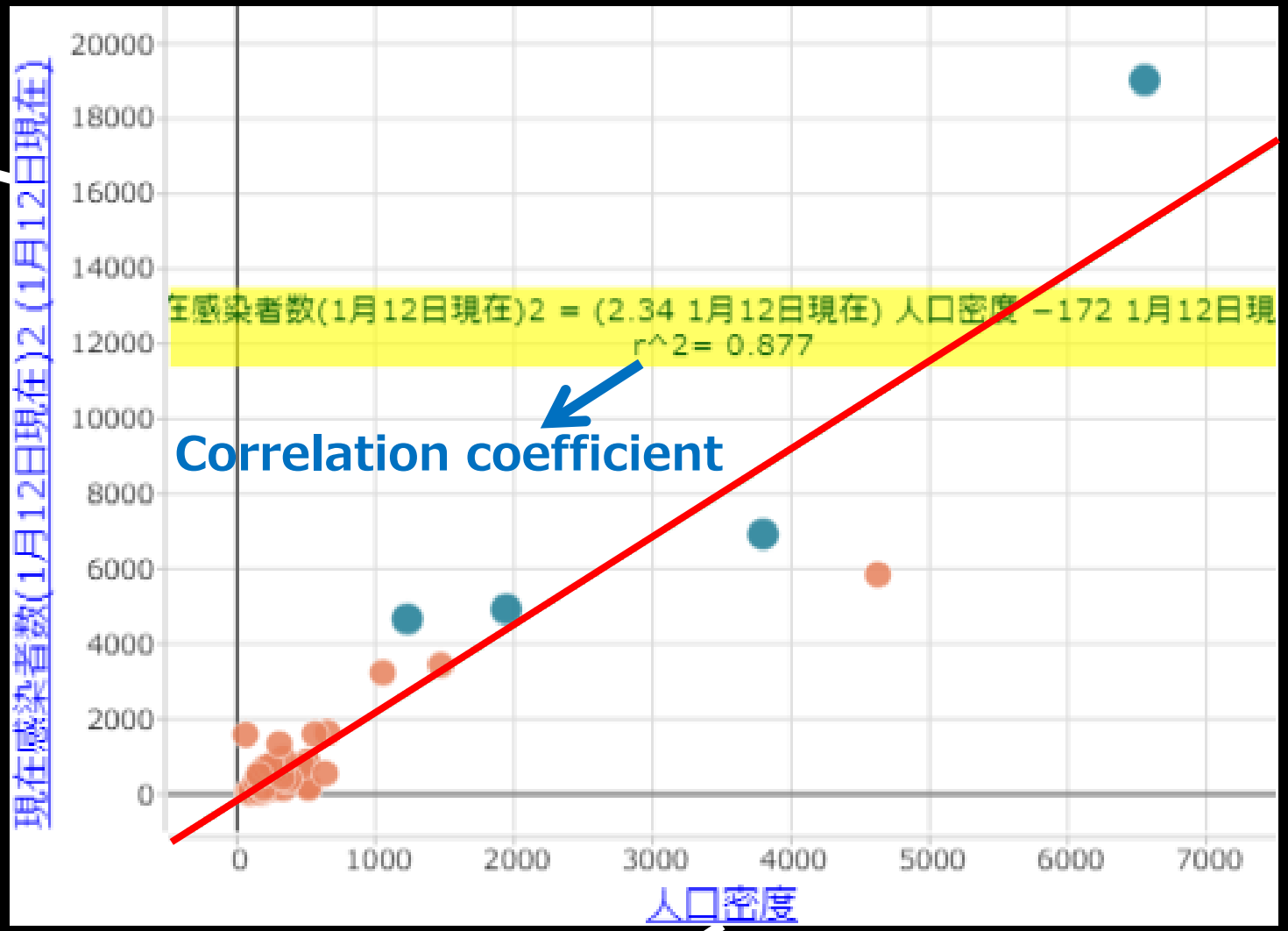
- Non-deterministic/stochastic models created and used by the PTs: dotplot, boxplot, scatterplot, and statistic (e.g., mean, median, quartiles, correlation coefficient).





Number of infected people in each of 47 prefectures

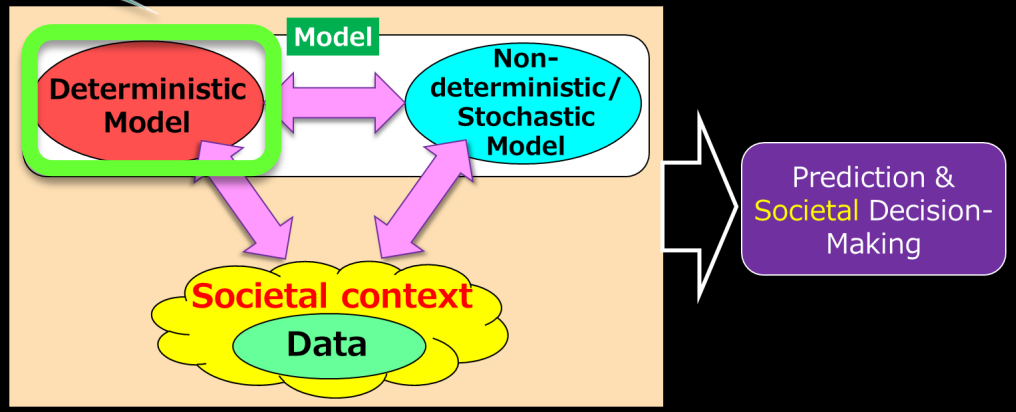
- This scatterplot can be considered a non-deterministic/stochastic model, given that it reflects a stochastic interpretation focusing on data variability.
- This regression line might also be considered a deterministic model, given that it reflects a deterministic interpretation focusing on exact numbers and causal explanation.

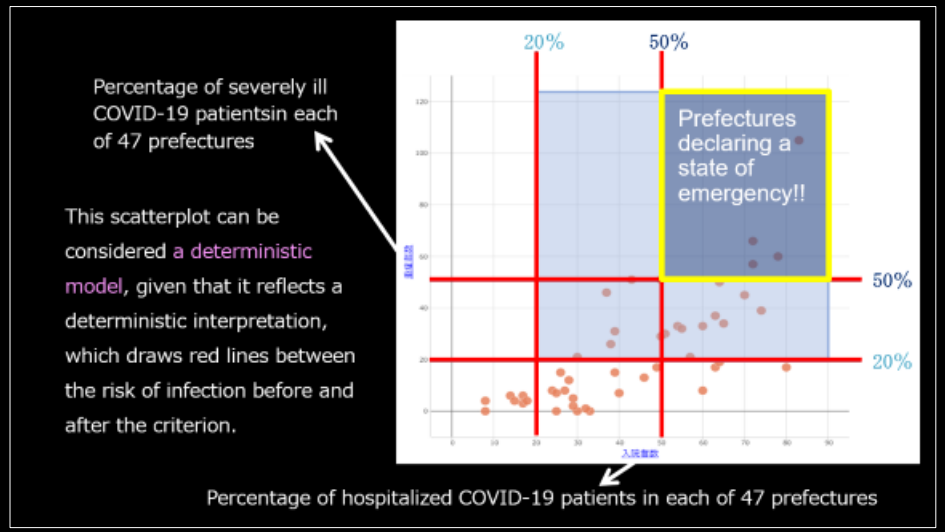


Population density in each of 47 prefectures

Pre-service teachers (PTs)' use of models given in the DDM framework (Kawakami & Saeki, 2024b)

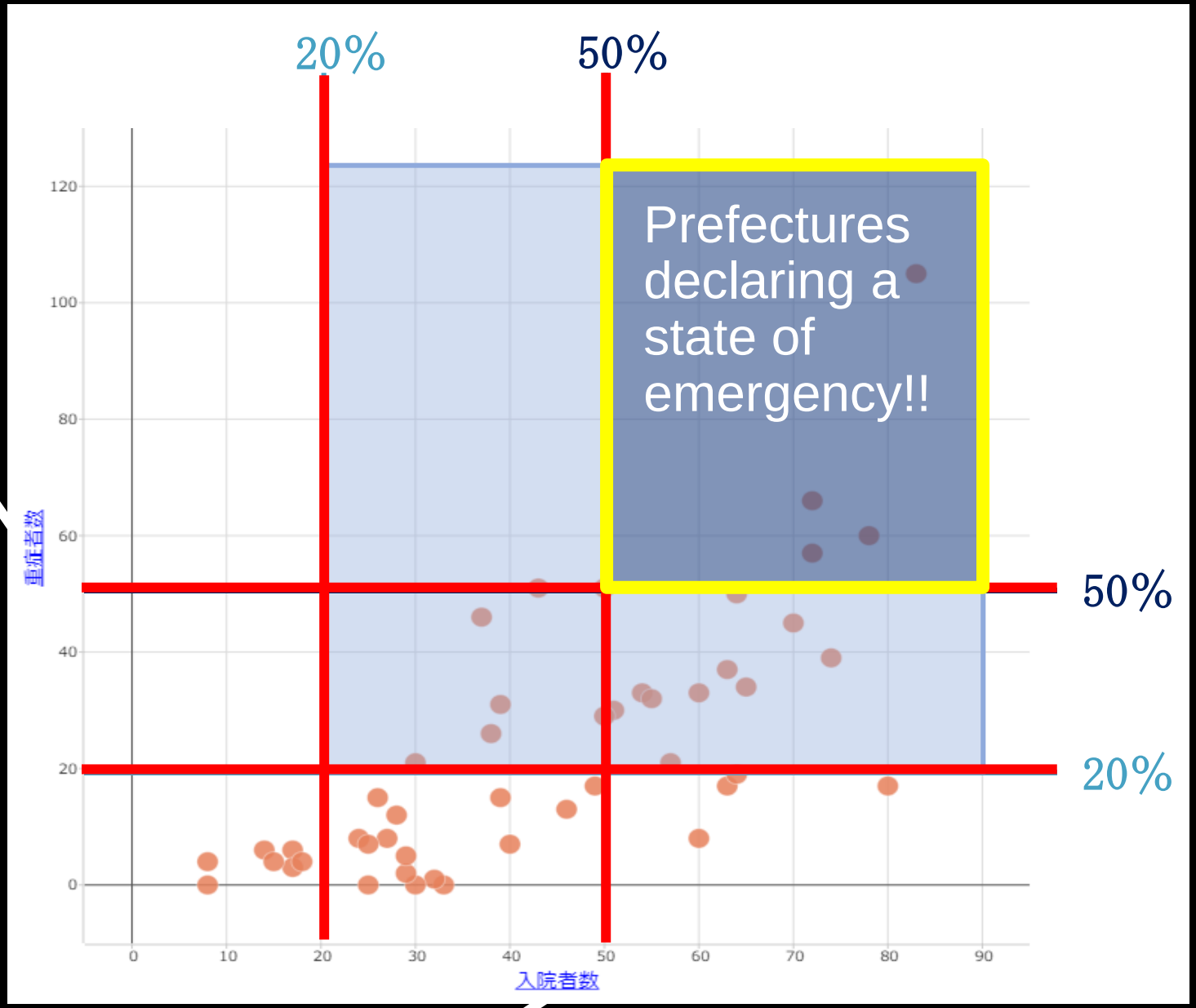
- Deterministic models created and used by the PTs: table, bar graph, line graph, dotplot, boxplot, scatterplot, regression line, and statistic (e.g., mean, median, quartiles).





Percentage of severely ill COVID-19 patientsin each of 47 prefectures

This scatterplot can be considered a deterministic model, given that it reflects a deterministic interpretation, which draws red lines between the risk of infection before and after the criterion.

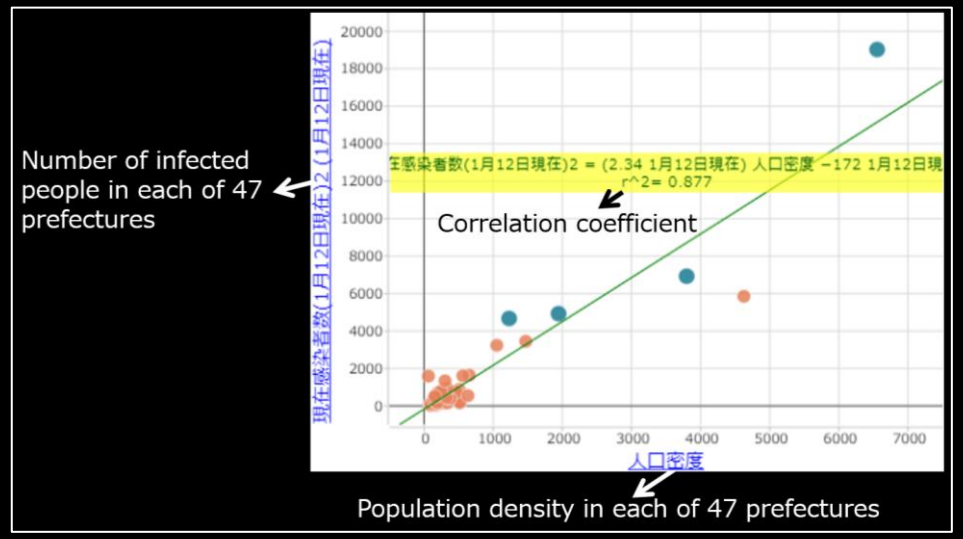


Percentage of hospitalized COVID-19 patients in each of 47 prefectures

How PTs used models (Kawakami & Saeki, 2024b)

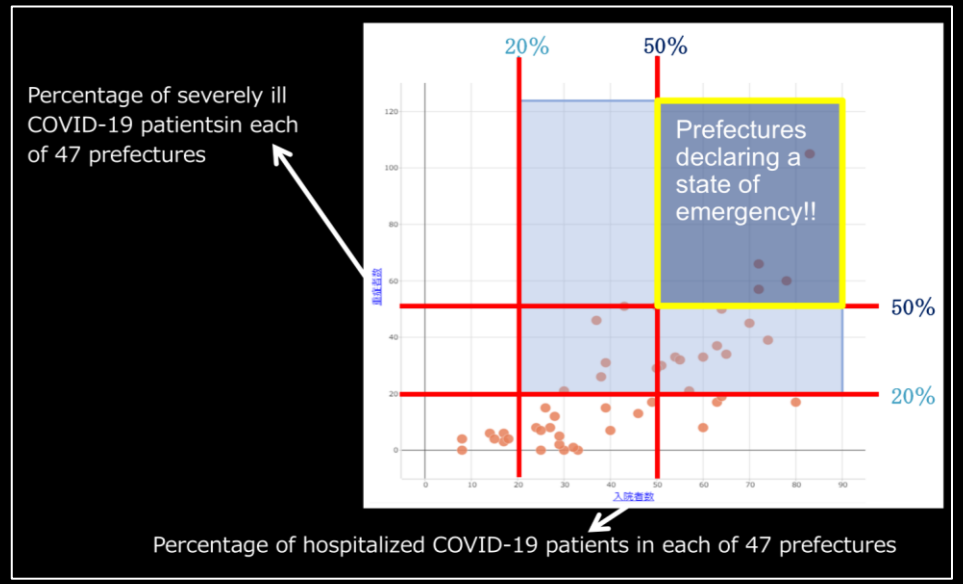
Non-deterministic/stochastic & deterministic models (e.g., scatterplot, regression line)

- Descriptive purpose to visualize the trends and/or variability of data on the current world

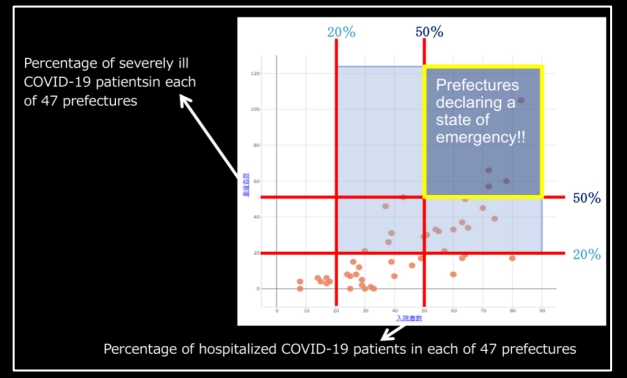


Deterministic model (e.g., scatterplot)

- Prescriptive/performative purpose to articulate data-informed societal decision-making and lead human action for a preferred world (Davis & Hersh, 1986; Niss, 2015; Skovsmose, 2024)



How PTs used models (Kawakami & Saeki, 2024b)



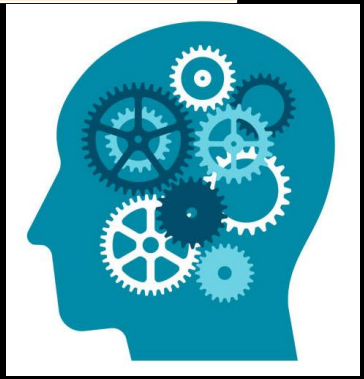
- All the participating PTs used deterministic models rather than non-deterministic/stochastic models as prescriptive/performative means in articulating their decision-making.

Why? ➢ This can be seen as a manifestation of PTs' ideology of certainty in mathematics education that *"mathematics can be applied everywhere and that its results are necessarily better than [those] achieved without mathematics"* (Borba and Skovsmose 1997, p. 18).

A PT's report that reflected on the importance of clarifying data-driven decision-making process

"I felt that I would not be able to explain decision-making well if I did not understand the process of my own data-driven reasoning. I realised that it was possible to incorporate the opinions of others and develop my reasoning by explaining the process rather than just the results. The decision-making process is different and complex for everyone and I was able to see the importance of explaining the process and the difficulty of decision-making while reflecting on own reasoning. So my awareness of the process of data-based reasoning increased."

The role of societal DDM in encouraging critical reflection



How can societal DDM contribute to broadly-defined data (science) educ. in primary & secondary schools?

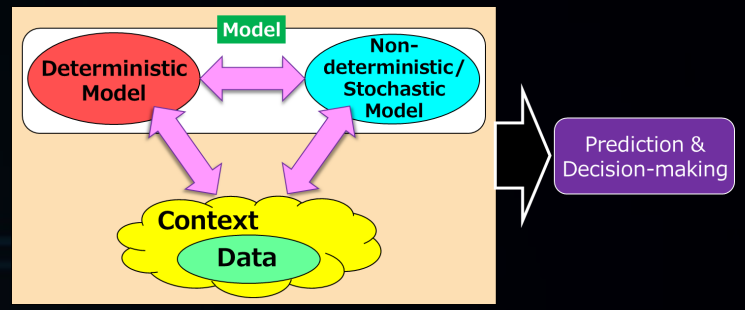
Mathematics, statistics, data science (and possibly AI) education

- Societal DDM can serve as an exemplary learning for future citizens as well as teachers to raise awareness and critical reflection on the prescriptive/performative power of data-informed deterministic and stochastic models and modelling process to reconstruct reality in society (i.e., O'Neil, 2016; Saltelli et al., 2020; Skovsmose, 2024).

(Saltelli et al., 2020, *Nature*)

5. Summary

Some potentials of DDM framework for broadly-defined data (science) educ. in primary & secondary schools



- A better understanding of learners' (students' & teachers') informal and diverse reasoning about data with mathematics and/or statistics
- A theoretical lens for clarifying, sharing, and discussing common perspectives between mathematics (mathematical modelling) education research and statistics/data science education research
- A practical lens for designing human-led modelling activities that leads to critical understanding of deterministic and stochastic models based on machine-led modelling using computational tools and AI

Some challenges & ways forward

- How should we ensure a balance between human-led and machine-led modelling in school and teacher education? (see also Fergusson & Pfannkuch, 2024)
- How do we develop DDM practices incorporating computational tools (simulation, programming, etc.) for primary and secondary schools?
- How do we develop and promote practices for primary and secondary students (also teachers) to learn about prescriptive/performative nature of data-informed models in broadly-defined data (science) education?

ありがとうございました!
Thank you very much for your attention!

The references cited in this talk are here!

Takashi Kawakami

Utsunomiya University, Japan

E-mail：t-kawakami@cc.utsunomiya-u.ac.jp

ORCID：https://orcid.org/0000-0002-0011-2155

ResearchGate：https://www.researchgate.net/profile/Takashi-Kawakami-6