Supplementary material

ODD protocol

In the following, the Overview, Design concepts, Details (ODD) protocol template (Railsback and Ayllón, 2020) has been adapted to provide a structured description of the DMW model. The purpose of the ODD is to provide transparency and information regarding the model's functionality.

Purpose and patterns

The DMW model is an aid designed for wastewater disease monitoring, aiming to model disease spread and mobility along with the dynamics of infected domestic wastewater (DW) across space and time based on the simulation of human behaviour. Applications include planning wastewater sampling (location and time) and translating sampling results to various disease cases. We address the question: How does modeling aid in understanding the spatiotemporal dynamics of infected domestic wastewater? The model is generic and can be easily applied to other diseases. The results can be generated at multiple temporal resolutions due to its event-based setup. The DMW reproduces realistic epidemic and DW patterns, and we have used the widely known patterns for the epidemic curves of the infected populations for the diseases. The model also reproduces pollutant patterns, such as total suspended solids (TSS). The spatiotemporal DW patterns are represented by day-to-day human activities at specific hours.

Entities, state variables, and scales

The DMW consists of the following entities: a) agents representing inhabitants and DW particles, b) the environment consisting of houses, neighbourhood blocks, schools, economic points, a sewer network, and a treatment plant. The information characterizing these entities is listed in Table 1, which has been adapted from DelaPaz-Ruíz et al. (DelaPaz-Ruíz, E.-W. Augustijn, *et al.*, 2023) to include the entities related to the disease submodel. The model runs dynamic time steps on the locality for the duration of the outbreak, where inhabitants' actions are based on discrete event simulations. Table 2 develops the attributes of the entities.

***Table 1**. Supplementary material. ODD entities of the DMW model. Modified from* (DelaPaz-Ruíz, E.-W. Augustijn, *et al.*, 2023).

Unit	**Variable**	**Description**	**Possible value**	**Data type**
Inhabitants	Age	Age category	18–24	String
	Study	Attend school	Yes	String
	Health	Health status	Susceptible	String
	Infection time	Date and time of infection	21-03-2022, 10:30:00	Time stamp
	Transmission type	Initial or secondary transmission	Secondary	String
	School level	School level	Highschool	String
	Work	Actively working	Yes	Boolean
	Gender	Inhabitant sex	Female	String
	CVEGEO	Block location	1302700010105004	Integer
	Ind ID	Individual ID	01750	Integer
	Location	Current location	Home, school, work	String
DW particle	DW type	Type of water appliance	Toilet stool	String
	Disease	Health condition	Infected	String
	Infection time	Date and time of infection	21-03-2022, 10:30:00	Time stamp
	DW quality	DW pollutants: COD and TSS	453	Integer
	CVEGEO	Block location	1302700010105004	Integer
	Ind id	Individual ID	01750	Integer
	DW speed	Traveling particle speed	1.8	Numeric
	Traveling path	Sewage network node	2020	Integer
Houses	House id	House ID	304	Integer
	CVEGEO	Block location	1302700010105	Integer

			004	
Blocks	CVEGEO	Block location	1302700010105004	Integer
Economic points	ID	Economic point ID	7996840	Integer
	CVEGEO	Block location	1302700010105004	Integer
	Avg. workers	Average workers	3, 5, 15	Integer
	School exists	School in this point	High school	String
Sewage network	Node	Connecting nodes	11	Integer
	Travel path	Node path to move DW particle	684	Integer
Treatment plant	Final station	The final destination of the traveling path	-	Integer

***Table 2**. Supplementary material. Description of attributes (or states variables).*

Unit	Variable	Change over time	Further description
Inhabitants	Age	Static	Differentiates calculation of pollutant loads and probabilities of water usage between age groups.
	Study	Static	Identifies which inhabitants attend schools.
	Health	Dynamic	Tracks health status changes during commuting.
	Infection time	Dynamic	Records the timestamp of health status changes according to the SIR model execution.
	Transmission type	Dynamic	Records the type of health status change according to the SIR model execution.
	School level	Static	Assigns inhabitants to their respective educational level locations (e.g., elementary school).
	Work	Static	Identifies which inhabitants are actively working.
	Gender	Static	Differentiates calculation of pollutant loads and probabilities of water usage by gender.
	CVEGEO	Static	Assigns inhabitants in their respective neighbourhood.
	Ind ID	Static	Differentiates and tracks events by individual ID.
	Location	Dynamic	Tracks inhabitants' location changes during commuting.
DW particle	DW type	Dynamic	Stores and differentiates the water appliance sources for the calculation of pollutant loads and events with viral concentrations (i.e., stool, and urine).
	Disease	Dynamic	Stores and differentiates the health status based on the inhabitant that generated the water appliance event (i.e., infected or recovered).
	Infection time	Dynamic	Stores the infection timestamp based on the infected inhabitant that generated the water appliance event.
	DW quality	Dynamic	Stores the pollutant load value. Calculated after

			postprocessing.
	CVEGEO	Static	Stores the CVEGEO location of the inhabitants in their respective neighborhoods.
	Ind ID	Static	Differentiates and track inhabitants' events by their ID.
	DW speed	Static	Defines the average movement speed of wastewater particles.
	Traveling path	Static	Guides wastewater particles according to the sewer direction.
Houses	House id	Static	Allocates inhabitant agents to households.
	CVEGEO	Static	Differentiates households by their neighbourhood block IDs.
Blocks	CVEGEO	Static	Differentiates the neighbourhood block IDs.
Economic points	ID	Static	Differentiates between economic points and commute inhabitants at school or work locations.
	CVEGEO	Static	To differentiate economic points between the neighbourhood block IDs.
	Avg. workers	Static	Defines the allowed number of inhabitants at the economic point.
	School exists	Static	Differentiates between school locations and the economic points.
Sewage network	Node	Static	Differentiates between the maintenance holes.
	Travel path	Static	Guides wastewater particles based on the sewer directions.
Treatment plant	Final station	Static	Identifies the final station for the motion of wastewater.

The DMW model facilitates simulation and representation of multiple spatial and temporal resolutions. The model uses the NetLogo time extension. The duration of the simulation is defined by the user which can be set with a specific time stamp. However, a range of two months is sufficient to represent the integration of the disease outbreak, mobility, and wastewater production when setting S (4), β (0.4). The input data shapefiles loaded in Netlogo defines the model's spatial scale.

Design concepts

Basic principles

The DMW model extends the domestic wastewater (DW) ABM(DelaPaz-Ruíz, E. Augustijn, *et al.*, 2023) by integrating a disease outbreak based on the Susceptible-Infected-Recovered (SIR) model. The DMW also integrates a population mobility model, which is essential for triggering the interactions between inhabitant agents and the spread of the disease in the locality. When inhabitant agents commute to school or work, they produce wastewater particle agents. Wastewater particles are produced by inhabitant agents via water appliances. Wastewater agents simulate the flow of sewage and store the health status of the inhabitant agents, allowing for the definition of the production of infected stool and the tracking of infected wastewater particles in time and space across the sewer network. Disease spread is activated when an infected inhabitant interacts with a susceptible inhabitant in the same location (e.g., school) and randomly infects the other inhabitants (see the disease model section). Inhabitants stop commuting when infected until they recover but continue producing wastewater particles reflecting the ongoing health status of these inhabitants.

Emergence

Emergence occurs through the simultaneous combination of the DMW submodels. Depending on the spatial configuration of the locality's environment and interactions between inhabitant agents, two main targeted spatiotemporal patterns emerge. The first is the epidemic curve pattern generated by the spread of the disease at specific locations among the inhabitant agents. The rules governing susceptible, infected, and recovered states are associated with this emergence. The second emerging pattern is related to the infected wastewater particle agents. This emergence is visualized as sewer flow where particles accumulate. The count of infected particles increases and decreases as the disease spreads. The relevant rules associated with this phenomenon include the flow motion in the sewer network, the disease spread that transmitting the health condition of inhabitants, and the prolonged prevalence of infected particles in the recovered population.

Interactions

Inhabitant agents interact with a random subset of other agents to facilitate disease spread. Two types of interaction groups exist: household members (other inhabitant agents living in the same house) and interactions at school or work. The 'S' parameter dictates the number of other agents with whom interaction takes place. The model does not simulate interactions between infected, susceptible, and recovered inhabitants indiscriminately; it restricts interactions to those between infected and susceptible inhabitant agents. Interactions are generated via random sampling. When infected and susceptible inhabitants are located at the sample location, a human-to-human interaction function $F(HH)$ activates, generating the infections.

The $F(HH)$ is linked to mobility, meaning that it is executed twice daily for all inhabitants who move or have the 'infected' status: once when inhabitants travel to school and work, and once when they return home. The interaction is defined by the 'S' parameter range, which indicates how many possible agents an infected agent has contact with resulting in infection. A transmission rate 'β' is used to define the probability of infecting a susceptible inhabitant. The range 'S' of agents is sampled randomly from the locations that an infected agent visits (household, school, workplace). When the infected and susceptible condition is met, the infected inhabitant is assumed to interact with a given number of other random inhabitants in the same location.

The NetLogo/Java script of the $F(HH)$ is as follows:

```
Let target one-of other inhabitants-here with[Disease_state="Infected"]
if random-float1<'β'[
let mylist.of.new.infections(range 'S'0-1)
    repeat one-of mylist.of.new.infections[
get-sick
settransmissionType"Secondary"]
]
```

The mathematical expression of $F(HH)$is as follows. Let:

- I be the set of inhabitants at the current location.
- $I_{\text{infected}} \subset I$ be the subset of infected inhabitants.
- $\beta_{\text{transmission}}$ be the transmission probability.
- S be the range of possible infections from an inhabitant, which can be expressed as a list.
- T be the transmission type, initially "Primary" but changing to "Secondary" upon transmission.

Step 1. Select a target:

$$target \in I_{\text{infected}} \tag{1}$$

Step 2. Probability check:

$$If\ Randomfloat \in [0,1) < \beta_{\text{transmission}},\ then \tag{2}$$

Step 3. List and repeat actions (create $mylist$, and iterate the actions for each element):

$$\begin{gathered} mylist = S \\ For\ randomly\ selected\ elements \\ \in mylist, \{Perform \\ get - sick: Set\ transmission\ \text{Secondary} \end{gathered} \tag{3}$$

In summary, the mathematical expression of $F(HH)$can be broken down into the following steps to capture the essence of the $F(HH)$ interaction in a mathematical framework: The selection of a random infected inhabitant set of current inhabitants. A subsequent check of whether a random number drawn from a distribution [0, 1) is less than the transmission probability. Then, if the probability check is satisfied, a range list is created and the infection process is repeated for one randomly selected element from this list, setting the transmission type to "Secondary".

Observation

The observation in the DMW records every agent event as the target variable of the retrieved patterns of each submodel. For every inhabitant agent, it records the execution timestamp (i.e., when they infect or move), location (house ID, school, work location), and health status (susceptible, infected, recovered). For every wastewater agent, it records the execution timestamp (e.g., the arrival at a maintenance hole), location (manholes and inhabitant IDs), and source descriptive information from the agent that generated the wastewater (e.g., such as use of a water appliance emitting urine or stool, and health status such as susceptible, infected, or recovered). The number of infected inhabitant agents per day is observed for the disease model. The count of inhabitants that remain at home during the outbreak evolution indicates mobility. The count of infected wastewater particles, including their arrival at a specific location in the sewer system (maintenance holes), is observed for the infected wastewater.

Stochasticity

The number of inhabitant agents per neighbourhood block is fixed; however, their distribution among households within the blocks is randomized. The inhabitants loaded into the ABM have individual census information, which assigns them to their respective school activities. The agents are randomly located at an economic point for work activities. Stochasticity is also present in wastewater variability, which fluctuates depending on the time of day and the presence of individuals in specific locations. Domestic water usage and the dynamics of emerging pollutant loads are modelled as probabilistic events to represent the inherent variability of DW loads.

Initialization

A simulation starts with loading the environments (locality, houses, economic points, schools, sewer pipes, maintenance holes, and treatment plant.). Then, the synthetic inhabitants generated via spatial microsimulation are loaded into the ABM and assigned to houses based on their location ID from the census data. Spatial microsimulation provides individual census data that allow the ABM to assign agents to their corresponding schools or economic location to represent work activities. At the start of the simulation, all inhabitant agents are susceptible. The disease submodel starts with five randomly infected individuals (inhabitant agents). The wastewater submodel includes the release of wastewater particles after initialization.

Input data

Table 3 lists the input data used in the DMW. The DMW uses geographical census data produced by the federal authorities of Mexico (INEGI, the National Institute of Statistic and Geography). There are two census datasets: one for economic activities (INEGI-DENUE, 2017) and another for the population (INEGI, 2020). The census data also include the geographical boundaries of the locality, neighbourhood blocks, and number of houses in each block. The household data were digitized using census information and satellite images.

The list of inhabitants was produced using spatial microsimulation (Lovelace and Dumont, 2018). Survey information related to the sewer network was digitized via a dedicated field data collection exercise in 2022, and included pipework, maintenance holes, and the treatment plant. This was supported by liaising with the local authorities to verify the network connectivity as much as possible. We used topographic data from a digital elevation model based on SRTM Tiles for digitization, which were acquired via the NASA Server QGIS plugin (Duester, 2021) and the Drinking Water, Sewerage and Sanitation Manual (CONAGUA, 2019) applicable to Mexico. Figure 1 shows the flow direction of the sewage.

***Table 3**. Supplementary material. Required input data for the DMW.*

Files	Notes	Types
Census: Locality	Polygon area	Shp

Census: Neighbourhood block	Multiple polygon areas	Shp
Census: Economic data	Multiple points	Shp
Census: Households	Multiple points	Shp
Census: List of inhabitants	Refers to the spatial microsimulation	CSV
Survey: Maintenance holes	Multiple points	Shp
Survey: Sewer network	Multiple lines	Shp
Survey: Treatment plant	Point	Shp

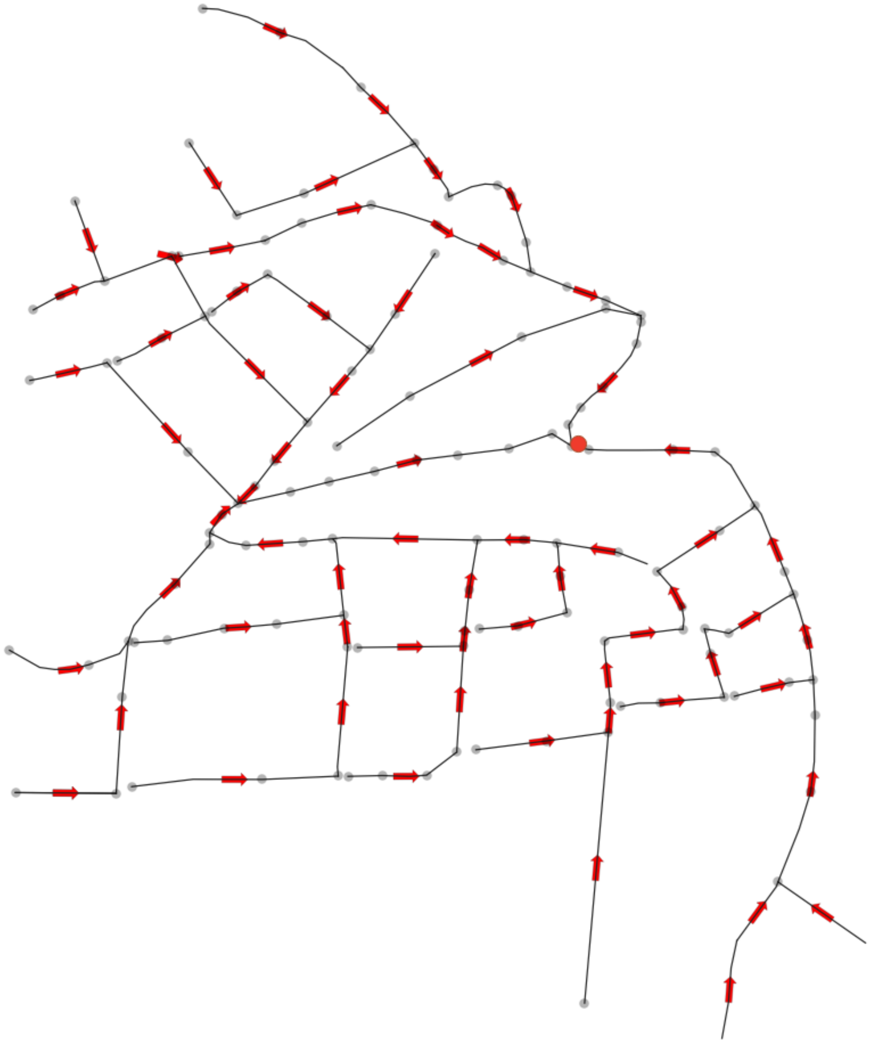

Figure 1. Supplementary material- sewage flow. Red arrows show the direction of the flow. The red point represents the WWTP. Lines represent the pipes. Grey dots are the maintenance holes.

Submodels

When the disease submodel executes, agents have a health state that can be either susceptible, infected, or recovered. At the start of the simulation, a number of initial infected cases (N) is generated. After the duration of infection of an inhabitant (D) has elapsed, the infected agent will recover(McMahon and Robb, 2020) and become immune until the end of the simulation. The DMW encapsulates the dynamics of how people commute between home, school, or work locations, reflecting the propagation of the disease throughout the population over time. The disease model includes an interaction component that determines the agents with which an agent has contact. Full interaction is assumed within the household, meaning that agents living in the same household can infect all other agents living in the house. In the cases of school and work, it is important to consider that individuals do not interact with all others present at the same location. When disease transmission requires close contact, we need to sample a subset of agents that can be infected, which is achieved using the Susceptible-Infected interaction *S* parameter. The '*S'* parameter indicates how many other agents an individual may contact when their locations overlap. The contacted agents are sampled randomly from the locations that an infected agent visits (i.e., household, school, or workplace).The disease outbreak is the first generated raw output, and its post-processing involves targeting the epidemic curves of the infected population and stool production at a daily resolution. The disease data includes the date (with daily timestamps); infected population; infected stools; and infected stools after the population recovers. Run ID to differentiate between the simulations. The disease raw data is retrieved from the actions of infected inhabitants and DW particle agents executed at any location inside the simulated environment (locality level).

When the mobility submodel executes in the DMW, inhabitant agents simulate mobility to represent inhabitant agents going to school or work and returning home based on working and studying hours, or staying at home, depending on their census data attributes. The modeling components and information linked to mobility are listed in Table 1, which are used to send inhabitants to specific school points and economic points (considered workplaces). For example, Table 1 shows an inhabitant high school agent; the ABM will send the agent to the high school's designated location. Since there are data limitations for mobility, mobility is simulated inside the limits of the case study. In other words, the local population does not leave the locality, and no individuals from the external population travel into the locality. The second raw dataset is the locality mobility and the household disease spread data, which produces a map series of infected inhabitants in households. The household disease data include date (with daily timestamps), household IDs, number of infected inhabitants and run IDs. It is retrieved at the end of each simulated day. The script also executes temporal and spatial aggregation that adds the infected inhabitants based on the household IDs. The following locality mobility data are recorded: date-time (with hourly timestamps), number of inhabitants at home, and run IDs to differentiate between the simulations. The raw mobility data are also retrieved from households, and the R script produces a smoothed plot to show the mobility pattern of households during the outbreak period from all simulations.

When wastewater executes in the DMW, each inhabitant agent will generate wastewater using toilets, water basins, showers, kitchens, and washing machines. Wastewater particles will be generated at the current location of the agent. When an inhabitant agent is infected, the wastewater particle agent will reflect the health status of the inhabitant. Wastewater moves via the sewer to the wastewater treatment plant. The DMW simulates the number of infected wastewater particle agents and the spatiotemporal DW loads of the infected stool (alternatively other DW loads can be analysed), representing the main output variables. The DMW allows for the tracking of infected wastewater by simulating the inhabitants' water appliance usage. Each inhabitant agent produces wastewater (hatching a DW particle agent) using water appliances. The DW production is based on distribution probabilities (DelaPaz-Ruíz, E.-W. Augustijn, *et al.*, 2023) that reflect typical daily sequences of water appliance usage for a given date and time. The execution probabilities of the agents actions also contain census-related information. The water appliances represent basins, kitchens, showers, toilets, and washing machines at their houses. Based on the literature, every single DW particle contains pollutant loads such as TSS. DW particle agents execute flow motion in the sewer, generating particles that move with an average velocity through the pipelines according to the network connectivity. The DMW model can provide the ID of each maintenance hole in the sewage and store the spatial ID data (see supplementary material) as a time series of pollutant loads, providing the DW with temporal variability. The infected wastewater is the third generated raw output, and its post-processing produces a map series of infected wastewater particles at maintenance holes at three different temporal resolutions (daily, hourly and every 6 minutes). The raw dataset includes date-time columns (with minute timestamps), wastewater particle IDs, water appliances (to identify the source of infected wastewater), the health statuses of inhabitants and run IDs. The third raw dataset is retrieved from the actions of the infected DW particle agents upon their arrival at a maintenance hole in the sewer (catchment level). The R script filters the data for a given time range and infected wastewater events. The script also performs temporal and spatial aggregation, adding the infected wastewater events based on the maintenance hole IDs for each target temporal resolution (6 minutes, 60 minutes and 24 hours) for 25 runs.

References

CONAGUA (2019) Manual de Agua Potable, Alcantarillado y Saneamiento. Volumen 20 Alcantarillado Sanitario, MAPAS: Manual de agua potable, alcantarillado y saneamiento. Available at: http://mapasconagua.net/libros/SGAPDS-1-15-Libro25.pdf.

DelaPaz-Ruíz, N., Augustijn, E.-W., et al. (2023) 'Modeling spatiotemporal domestic wastewater variability: Implications for measuring treatment efficiency', Journal of Environmental Management, 351(Sustainable Solid and Liquid Waste Management), p. 12. Available at: https://doi.org/10.1016/j.jenvman.2023.119680.

DelaPaz-Ruíz, N., Augustijn, E., et al. (2023) 'Software. Reproducible results. Modeling spatiotemporal domestic wastewater variability: Implications for measuring treatment efficiency.' Zenodo. Available at: https://doi.org/10.5281/zenodo.10242567.

Duester, H. (2021) 'SRTM-Downloader: Plugin for download of SRTM Tiles from NASA'. Available at: https://github.com/hdus/SRTM-Downloader/wiki (Accessed: 7 October 2022).

INEGI-DENUE (2017) Directorio Estadístico Nacional de Unidades Económicas (DENUE) 2017. Instituto Nacional de Estadística y Geografía. INEGI. Available at: http://www.beta.inegi.org.mx/app/descarga/?ti=6.

INEGI (2020) Censo de Población y Vivienda 2020, Censos y conteos. Available at: https://www.inegi.org.mx/programas/ccpv/2020/ (Accessed: 7 October 2022).

Lovelace, R. and Dumont, M. (2018) Spatial Microsimulation with R. Website ed. CRC Press. Available at: https://doi.org/10.1201/9781315381640.

McMahon, A. and Robb, N.C. (2020) 'Reinfection with SARS-CoV-2: Discrete SIR (Susceptible, Infected, Recovered) modeling using empirical infection data', JMIR Public Health and Surveillance, 6(4), p. e21168. Available at: https://doi.org/10.2196/21168.

Railsback, S. and Ayllón, D. (2020) 'Supplementary file S1 to: Grimm, V. et al. (2020) "The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism"', JASSS-Journal of Artificial

Societies and Social Simulation [Preprint]. Available at: https://doi.org/10.18564/jasss.4259.