Less than two days of water quality sampling at local beaches may be all that's needed to reduce illnesses among millions of beachgoers every year due to contaminated water, according to new Stanford research. The study, published in Environmental Science & Technology, presents a modeling framework that dependably predicts water quality at beaches after only a day or two of frequent water sampling. The approach, tested in California, could be used to keep tabs on otherwise unmonitored coastal areas, which is key to protecting the well-being of beachgoers and thriving ocean economies worldwide.

"This work combines knowledge of microbiology, coastal processes and data science to produce a tool to effectively manage one of our most precious resources and protect human health," said senior author Alexandria Boehm, a Stanford professor of civil and environmental engineering.

Measuring concentrations of fecal indicator bacteria (FIB) – which denote the presence of fecal matter and can lead to unsafe water conditions—at beaches ensures the health and safety of the public. While all ocean water contains some degree of pathogens, such as bacteria or viruses, they're typically diluted to harmless concentrations. However, changes in rainfall, water temperature, wind, runoff, boating waste, storm sewer overflow, proximity to waste treatment plants, animals and waterfowl can lead to an influx of water contamination. Exposure to these contaminants can cause many ailments, including respiratory diseases and gastrointestinal illnesses, along with skin, eye and ear infections to swimmers.

Protecting coastal waters and the people that use them remains essential for much of California's 840 miles of coastline. Over 150 million people swim, surf, dive and play at one of the state's 450 beaches annually, generating over $10 billion in revenue. According to the California State Water Resources Control Board, health agencies across 17 counties, publicly owned sewage treatment plants, environmental groups and several citizen-science groups perform water sampling across the state. However, not all waters are routinely checked due to accessibility issues, budget resource constraints or the season, despite their use by the public.

Another obstacle to safeguarding public health lies in the lag time between sampling and results—up to two-days—leading beach managers to make decisions reflecting past water quality conditions. When monitored waters contain high levels of bacteria and pose a health risk, beach managers post warning signs or close beaches. The delay in current testing methods could unknowingly expose swimmers to unhealthy waters.

To overcome these limitations, the researchers combined water sampling and environmental data with machine learning methods to accurately forecast water quality. While predictive water quality models aren't new, they have generally required historical data spanning several years to be developed.

The team used water samples collected at
10-minute intervals over a relatively brief timeframe of one to two days at beaches in Santa Cruz, Monterey and Huntington Beach. Among the three sites, 244 samples were measured for FIB concentrations and marked as above or below the acceptable level deemed safe by the state. The researchers then collected meteorological data such as air temperature, solar radiation and wind speed along with oceanographic data including tide level, wave heights and water temperature (all factors influencing FIB concentrations) over the same timeframe.

Using the high-frequency water quality data and machine learning methods, they trained computer models to accurately predict FIB concentrations at all three beaches. The researchers found hourly water sampling for 24 hours straight—capturing an entire tidal and solar cycle—proved enough for reliable results. Feeding the framework meteorological and tidal data from longer time periods resulted in future water quality predictions that were dependable for at least an entire season.

"These results are really empowering for communities who want to know what's going on with water quality at their beach," Searcy said. "With some resources to get started and a day of sampling, these communities could collect the data needed to initiate their own water quality modeling systems."

The framework code, which is publicly accessible, could also be developed for accurate predictions of other contaminants such as harmful algae, metals and nutrients known to wreak havoc on local waters. The researchers point out that more analysis is needed to better determine the exact timeframe these models remain accurate and note that continually assessing and retraining the models remains a best practice for accurate predictions.
