The Influence of Outdoor Conditions on Indoor Air Quality: Case Study of Norwegian Schools

Iselin Ørbek Eide[1] Christer Eskedal[2], Kai Gustavsen[3], Kent Hart[3], Azimil Gani Alam [1], Guangyu Cao[1]

1 Department of Energy and Process Engineering (EPT)
Norwegian University of Science and Technology (NTNU),
Kolbjørn Hejes v 1B
7034 Trondheim. Norway

2 N3
Rudssletta 60,
1351 Rud, *Norway*

3 Norges Astma- og Allergiforbund
Gnr 14 Bnr 4
7039 Trondheim, Norway

corresponding email: Iselin.o.eide@ntnu.no

ABSTRACT

The project aims to investigate the degree of influence that outdoor conditions may have on the indoor environment in Norwegian schools. It also aims to ascertain whether it is possible to use outdoor parameters such as particulate matter, relative humidity, and air temperature, along with indoor parameters including CO2, relative humidity, and air temperature, to predict indoor particulate matter values. The outdoor data was gathered from various weather stations near the schools, while the indoor data was provided by N3, who collected it using sensors within the schools. To predict indoor particulate matter values, a machine learning algorithm, random forest regression, was employed. The project's findings highlight the significant impact of outdoor conditions on indoor environments. These encompass a wide spectrum, ranging from local weather effects, like cold temperatures leading to pollutant accumulation at lower altitudes, to remote occurrences, such as Sahara sands traveling thousands of kilometres by wind to Norway. Variations in the correlation between indoor and outdoor PM10 values across different locations and classrooms suggest potential diverse sources of particulate matter, seasonal effects on indoor air quality, or disparities in ventilation systems and cleaning procedures. These initial findings were further investigated by a random forest regression algorithm in machine learning. This approach incorporated diverse outdoor and indoor parameters to assess feature importance and forecast indoor PM10 levels, resulting in robust models with achieved R-squared values reaching 0.92. In January, outdoor temperature emerged as the primary influential factor, followed closely by outdoor PM2.5 values and indoor relative humidity. In contrast, September emphasized indoor relative humidity as the most significant influence. Notably, indoor CO2 demonstrated a consistent level of influence in both January and September, likely due to consistent student activity. The robustness of the random forest models and their close alignment with actual PM10 values suggest a strong potential for establishing future cleaning procedures based on predictive models.

KEYWORDS

Indoor Environment - Particulate Matter - Correlation Particulate Matter

1 INTRODUCTION

In recent years, the quality of Indoor Environmental Quality (IEQ) has emerged as a critical concern impacting health, well-being, and productivity on a global scale. The World Health Organization has highlighted the significance of IEQ, pointing out its direct correlation with public health outcomes across various settings (World Health Organization, 2010). Amidst this global context, the aspect of Indoor Air Quality in educational institutions, particularly elementary schools, demands special attention. Studies, including the 2019 National

Monitoring of Working Environment and Health (NOA) conducted by the National Institute of Occupational Health, reveal that 35 percent of elementary school teachers in Norway report poor IAQ due to inadequate ventilation, underscoring a widespread issue within educational settings.
The urgency to address IAQ is further magnified by the broader challenges posed by deteriorating outdoor air quality (Fenger, 2009). Pollution and environmental degradation exacerbate indoor air problems, creating a vicious cycle that impacts vulnerable populations, such as children in schools, the most. This interconnection between outdoor and indoor air quality underscores the necessity for comprehensive solutions that address both facets to ensure healthy living and learning environments.
Following the pandemic, there has been a global call from researchers for an indoor climate revolution, emphasizing the need for improved IEQ in schools to safeguard and enhance the learning experience (Eichholtz, 2024). The indoor environment plays a pivotal role in the health, well-being, and productivity of occupants, making it imperative to understand and optimize the factors influencing IAQ.
This study aims to investigate the feature importance of both indoor and outdoor parameters on indoor Particulate Matter (PM) values, uncovering the causes behind their significant impact on IAQ. Additionally, it seeks to develop a predictive model for indoor PM values using a random forest machine learning model, contributing to the body of knowledge on effective IEQ management and offering actionable insights for improving air quality in schools.

1.1 Scope and Delimitation

Due to time constraints and the scope of the assignment, certain simplifications had to be made. The analysis was restricted to specific months instead of the entire year to save time, and the data solely pertains to the year 2023. While this renders the analysis more vulnerable to yearly fluctuations, it was deemed sufficiently robust to validate the results.
Various data discrepancies emerged throughout the project. Instances of missing data arose due to sensors being offline, damaged, or due to other unidentified reasons. Consequently, this led to less accurate results and some data sets remained uncalculated.
Notably, not all outdoor sensors were positioned on the school grounds, thereby introducing some inaccuracies in the results. The outdoor data slightly differs from the actual outdoor conditions on the school grounds. Nevertheless, the sensors were placed close, and the selected schools were chosen because of their proximity to the weather stations, aiming to mitigate significant impacts on the analysis.

2 METHODOLOGY

2.1 Seasonal Investigation Parameters

Due to the vast amount of data and the limited time and scope of the assignment, specific time periods of the year were selected for investigation rather than examining the entire year.

The objective was to explore time segments from various seasons to compare their respective impacts on the indoor environment. Additionally, it was crucial to choose months when the school was in full operation. Consequently, the summer months of June, July, and August were not feasible due to summer vacation. Similarly, April was affected by the Easter break, and May had various national holidays. As a result, the months chosen for further investigation were January, representing the winter season, March for spring, and September for fall. Due to unforeseen issues with missing data from March in several schools, the machine learning model could not produce reliable predictions for this month.

2.2 Data Collection

The indoor data for the analysis were sourced from the N3smart sensors installed within the selected rooms, while the outdoor data originated from various weather stations located near the schools. Specifically, the outdoor PM10 and PM2.5 values were gathered from weather stations affiliated with NILU, an independent nonprofit research institution. The temperature and relative humidity data were obtained from the Norwegian Climate Service Centre (KSS), which serves as a hub for climate and hydrological data, facilitated through a collaboration between the Meteorological Institute, the Norwegian Water Resources and Energy Directorate, the Mapping Authority, NORCE, and the Bjerknes Centre.

Several different schools in Norway were chosen for the analysis, partly due to their proximity to weather stations. The schools are presented in Table 1.

Table 1: Selected schools and rooms

School Name	Location	Construction year	Room 1	Room 2
Stabekk	Bærum	2004	Ø203	V201
Åsveien	Trondheim	2015	Base 4	Locker room
Åsenhagen	Lillestrøm	2000	D-150	7D
Kjeller	Lillestrøm	2010	1020	1045
Høvik	Bærum	2013	3139	1007

The sensors that collected data inside the schools were provided by N3. These sensors gather data concerning CO_2, air temperature, relative humidity, and PM10. The frequency at which new data is measured can be adjusted, but for this project, it was set to every 2 minutes. These sensors are positioned in the middle of the classrooms on the wall opposite the door.

Various weather stations were used to source data for the analysis, and different weather stations had to be used for the same school due to the particulate matter values and the temperature and relative humidity values not being recorded at the same stations.
The weather stations used for PM_{10}, $PM_{2.5}$, relative humidity, and temperature can be seen in Table 2.

Table 2: Weather stations used for PM-values, relative humidity, and temperature. Distance from school in parentheses.

School	PM Weather station	Temp and RH Weather station
Stabekk Primary School	Bekkestua weather station (700 m)	Skriverberget weather station (2 km)
Åsveien Primary School	Åsveien skole weather station (50 m)	Saupstad weather station (3 km)
Åsenhagen Primary School	Vollaparken øst weather station (2 km)	Skedsmo-Hellerud weather station (700 m)
Høvik Primary School	E18 Høvik kirke weather station (500 m)	Skriverberget weather station (2,5 km)
Kjeller Primary School	Vigernes Weather station (3km)	Kjeller Weather station (300 m)

2.3 Machine Learning

Machine learning has become a proven method to be utilized in predicting indoor air quality (Wei, 2019). In this study, supervised ML with random forest algorithm was utilized to predict indoor PM10 values based on eight parameters. These parameters were employed to train the model and determine which had the greatest impact on indoor PM10 values. Recorded indoor PM10 values were used as training data for the model. The input parameters are listed in Table 3. Additionally, decision tree and linear regression models were employed, but the random forest model consistently yielded more accurate results and were therefore used for the results.

Table 3: Input parameters

Indoor Values	Outdoor Values
Relative humidity	Relative humidity
Temperature	Temperature
PM_{10}	PM_{10}
CO_2	$PM_{2.5}$

3 RESULTS

3.1 Outdoor Parameters

Figure 1 and Figure 2 display the outdoor PM10 values for all schools in January and September. The graph is divided into zones in compliance with TEK17s criteria for zoning in the planning of activities or construction. September exhibits lower values of PM10 than January. However, there is a noticeable increase for all schools between the 6th and 12th of September.

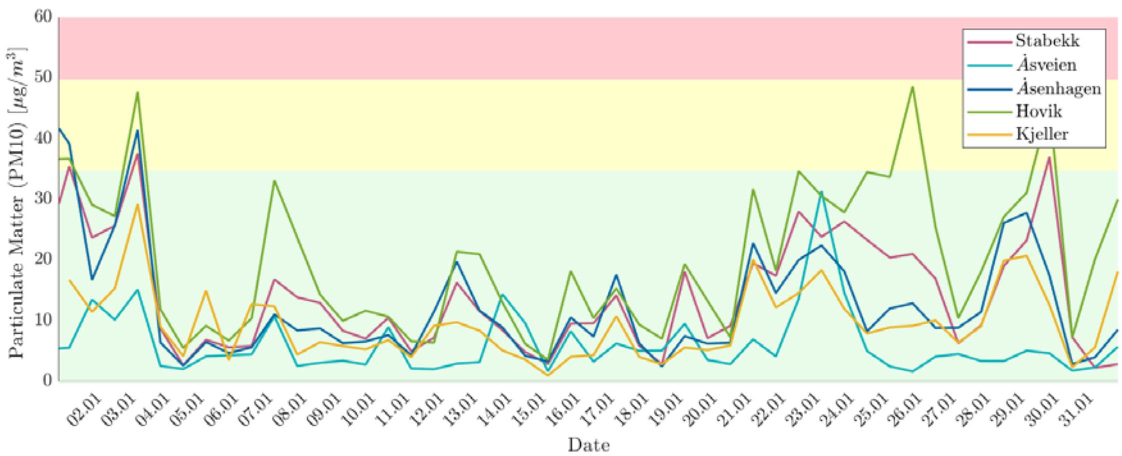

Figure 1: PM10 values outdoor for all schools, January.

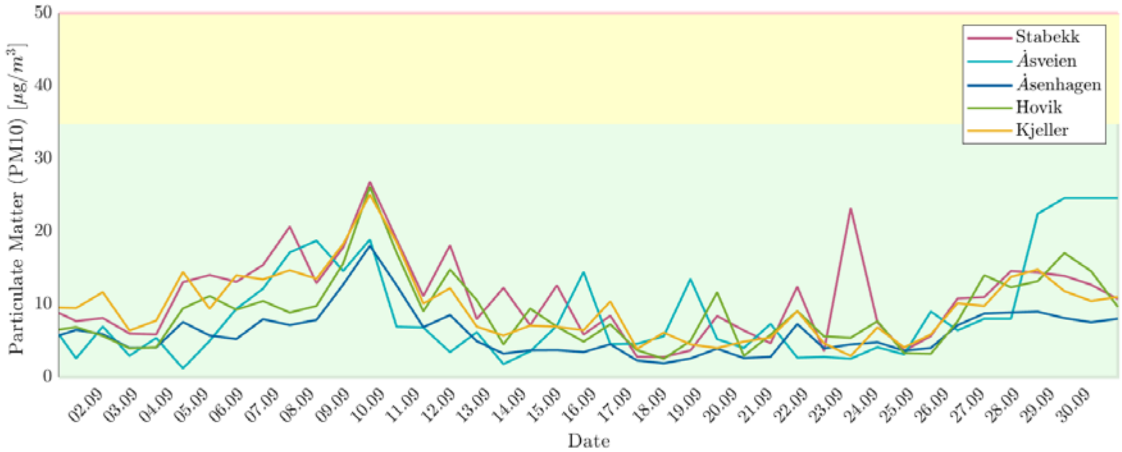

Figure 2: PM10 values outdoor for all schools, September.

Figure 3 and Figure 4 show the outdoor PM2.5 values for all schools. Both graphs follow the same trend as the PM10 graphs, albeit slightly smaller, which is natural since the PM10 values also include the values for PM2.5. The zones set by TEK17 contain information solely regarding PM10. However, since PM10 includes smaller particulate matter, it remains applicable to the PM2.5 graphs.

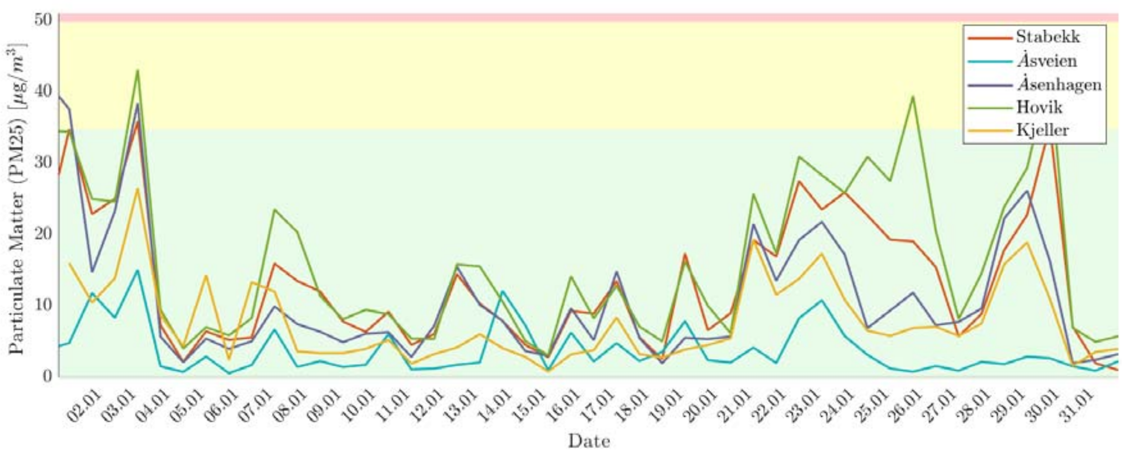

Figure 3: PM2.5 values outdoor for all schools, January.

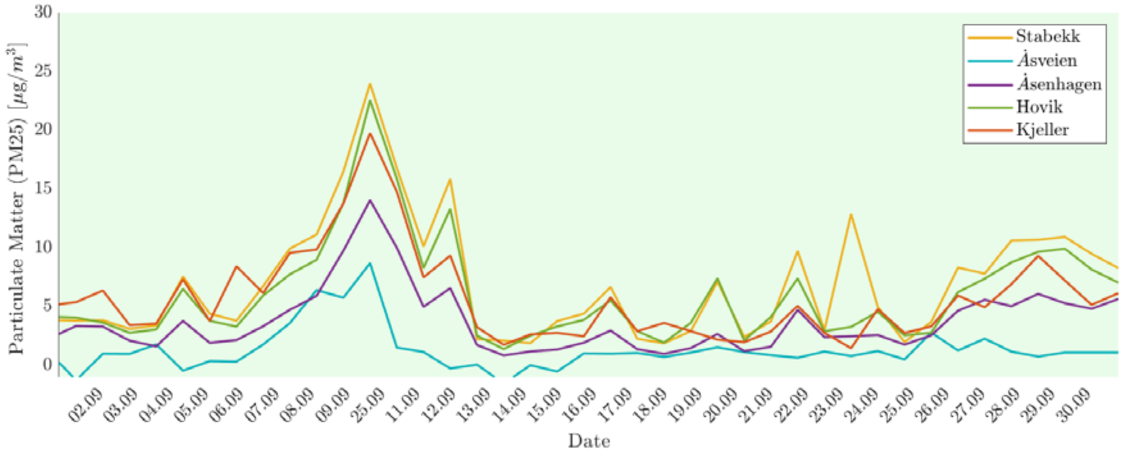

Figure 4: PM2.5 values outdoor for all schools, September.

Naturally, September was much warmer than January. Åsveien School exhibits the largest difference in temperature from the others, being situated much further north, in Trondheim. The relative humidity values from Stabekk and Høvik show unusually stable readings for a significant period in January, which might indicate sensor errors. Åsveien has some missing values for September. The relative humidity stays between 50 percent and 100 percent for both January and September, with Åsveien showing lower values than the other schools in January.

3.2 Correlation

The correlation between indoor and outdoor PM10 values is presented in Table 4. The correlation is calculated for the entire day and separately for workdays when ventilation is active. Some calculations involve incomplete outdoor or indoor data, leading to skewed results. Due to problems with several of the data sets for March, the month of March was removed from all the results.

Table 4: Correlation between indoor and outdoor PM10 values

School	Month	Sensor	Corr.	Corr. Daytime	Comment
Stabekk	January	Ø203	0.70	0.75	
		V201	0.71	0.79	
	September	Ø203	0.31	0.23	
		V201	0.36	0.29	
Åsveien	January	Sensor 4	0.35	0.40	
		Sensor 5	0.43	0.40	
	September	Sensor 4	0.49	0.44	
		Sensor 5	0.41	0.24	
Åsenhagen	January	D-150	0.59	0.52	Missing parts of outdoor data
		D7	-0.15	-0.13	Missing parts of outdoor data
	September	D-150	0.79	0.78	
		D7	0.75	0.69	
Kjeller	January	1020	0.29	0.42	Missing parts of indoor data
		1045	0.41	0.25	Missing parts of indoor data
	September	1020	0.52	0.47	
		1045	0.55	0.46	
Høvik	January	3139	0.52	0.50	
		1007	0.79	0.66	

September	3139	0.43	0.44
	1007	0.59	0.48

3.3 Predictive Patterns

Predictive patterns for the indoor PM10 values and what parameter that had the most significant impact were generated using a random forest model (regression). The Coefficient of variation (CV) and R^2 values for the models can be seen in Table 5Table 5. Most of the models were strong or very strong, capturing a substantial amount of the variability. The model only uses data from when the ventilation was going during the day.

Table 5: CV and R2 values for the predictive model

	January		September	
School/Room	**R^2**	**CV**	**R^2**	**CV**
Stabekk Ø203	0.803	0.803	0.813	0.814
Stabekk V201	0.812	0.813	0.743	0.743
Åsveien base 4	0.802	0.802	0.730	0.730
Åsveien locker room	0.839	0.845	0.665	0.669
Åsenhagen D-150	X	X	0.890	0.890
Åsenhagen 7D	X	X	0.887	0.888
Kjeller 1020	X	X	0.796	0.800
Kjeller 1045	X	X	0.789	0.789
Høvik 3139	0.783	0.786	0.634	0.634
Høvik 1007	0.801	0.803	0.922	0.923

Figure 5 shows which feature had the largest impact on the indoor PM10 values for January 2023. The number of schools is reduced due to missing data from Åsenhagen and Kjeller. The most influential feature in January was the outdoor temperature, closely followed by outdoor PM2.5 values and indoor relative humidity.

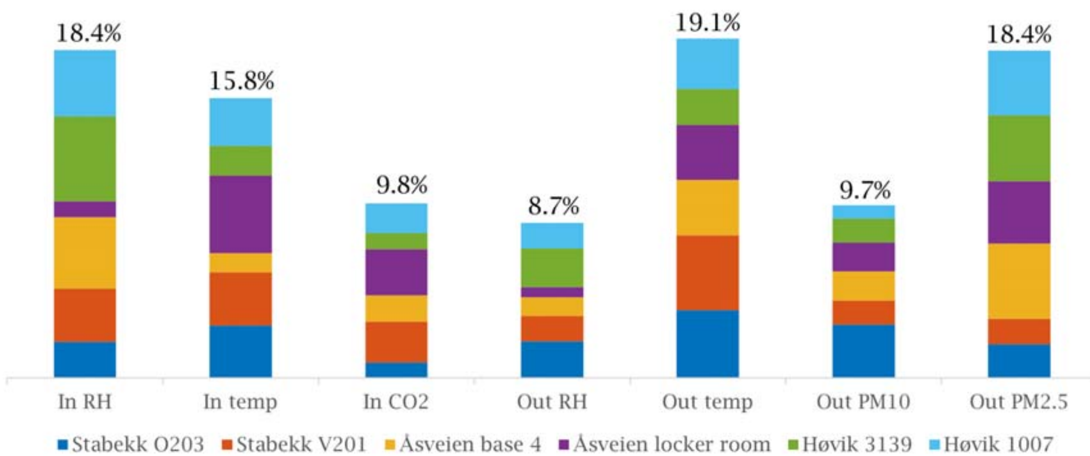

Figure 5: Feature importance, January

Figure 6 displays which feature had the most impact in each room, aiding in the distinction between room-specific differences.

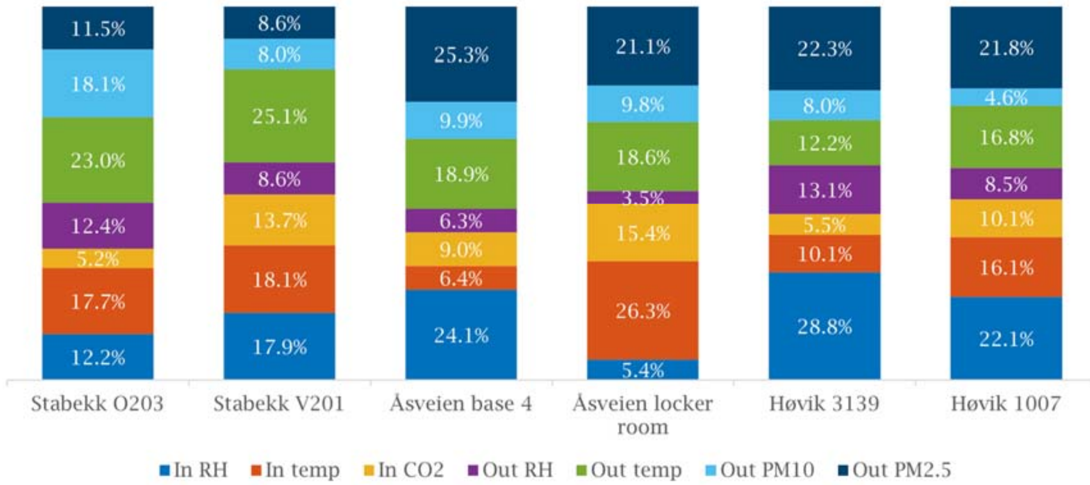

Figure 6: Feature importance for each room, January

Figure 7 shows which feature had the largest impact on the PM10 values for September 2023. It can be observed that indoor relative humidity had the most significant overall impact, albeit with some variations across different rooms.

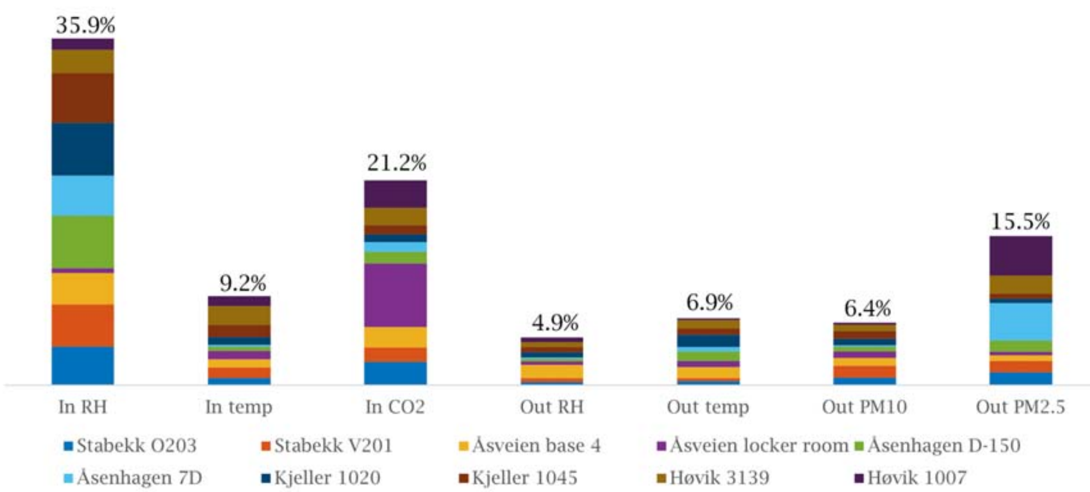

Figure 7: Feature importance, September

Figure 8 displays which feature had the most impact in each room, facilitating a clearer comparison between the rooms.

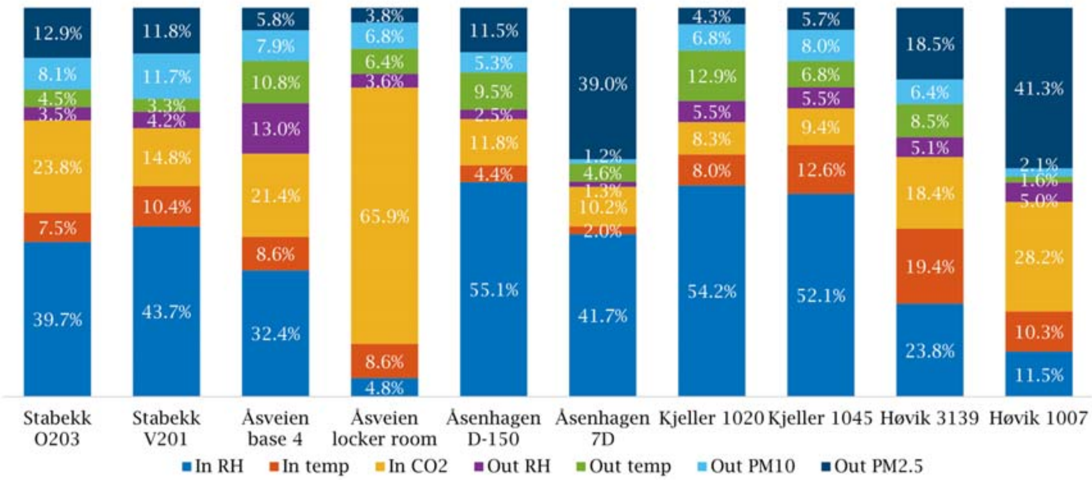

Figure 8: Feature importance for each room, September

4 DISCUSSION

4.1 Outdoor Parameters

The outdoor data for PM10and PM2.5 indicates that Stabekk and Høvik primary schools have the highest values of PM overall. Both schools are situated near the European route E18, the main road leading into the capital, Oslo. It is reasonable to assume that the elevated values are due to road dust. The PM values are also generally higher for January than for September. This results from the concentration of outdoor pollutants in the lower atmosphere due to temperature inversions. Typically occurring in cold weather, the ascent of warm air to the upper atmosphere creates a layer that confines colder air below it. Consequently, pollutants accumulate at lower altitudes.

All schools experienced a distinct increase in particulate matter between the 6th and 12th of September. This increase was caused by a storm in the Sahara Desert. The dust was carried by the wind all the way from the Sahara Desert to Bærum, Lillestrøm, and even Trondheim, resulting in elevated dust levels for all the schools.
The outdoor temperature and relative humidity for Stabekk, Høvik, Åsenhagen, and Kjeller were quite similar for both January and September. All four schools are situated close to each other, with Høvik and Åsenhagen being the farthest apart at 35 km. Therefore, it is expected for them to have similar air temperatures and relative humidity. Åsveien, located in Trondheim, showed the largest difference from the others, especially regarding relative humidity, which was significantly lower in January.

4.2 Correlation

The outdoor and indoor PM10 values exhibit varying degrees of correlation, as displayed in Table 4. The lowest correlation, 0.23, was observed for Stabekk in September, while the highest, 0.79, was noted for Stabekk in January.
In September, Stabekk and Høvik have a lower correlation compared to January, whereas Åsenhagen and Kjeller exhibit a higher correlation in September than in January. Åsveien demonstrates the highest correlation in September for one room and in January for the other room.

These correlations suggest a notable discrepancy in the relationship between outdoor and indoor PM10 concentrations. To further investigate the matter regarding outdoor factors a predictive machine learning model was utilized.

4.3 Predictive Patterns

The analysis utilized a random forest regression algorithm in the machine learning model. Most models generated by this algorithm demonstrated strength, capturing a substantial amount of variability. The model exclusively utilized data collected during operational ventilation hours. The corresponding CV and R^2 values for the model are presented in Table 5. Among these models, those for Åsenhagen and Høvik exhibited the highest strength, boasting R^2 values as high as 0.89 and 0.92, respectively.

In January, the most influential feature was outdoor temperature, closely followed by outdoor PM2.5 values and indoor relative humidity. Indoor temperature also exhibited significance, while indoor CO_2 values, outdoor PM10, and relative humidity values played minor roles.
In September, the most influential feature was indoor relative humidity. Indoor CO_2 exhibited a slightly larger level of influence compared to January. PM2.5played a minor role, while indoor temperature, outdoor temperature, PM10, and relative humidity had minimal importance.

The CO_2 values exhibit a similar level of influence on indoor PM10 values for both January and September, slightly higher in September. This influence is likely due to consistent student activity, which remains constant regardless of the seasons.
For January, the outdoor temperature had the biggest impact on the indoor PM10 values followed by the PM2.5 values. These two parameters can be related. The low temperatures make the pollutants accumulate at lower altitudes. When there is a higher concentration of PM2.5 they will naturally have a higher impact on the indoor values.
The indoor relative humidity had a considerable importance for both January and September. There can be several reasons for this. In lower humidity the static electricity in the dust particles can increase, causing dust particles to repel each other and remain airborne longer. At the same time, higher humidity can reduce static charges, encouraging particles to clump and settle faster.
Higher humidity can also cause the dust particles to absorb moisture and clump together, making them heavier and settle on the floor or other surfaces.

5 CONCLUSIONS

The project aimed to determine the degree of influence that outdoor conditions have on the indoor environment in Norwegian schools. Additionally, it sought to ascertain whether using outdoor parameters would enable the prediction of indoor particulate matter values.
From the project findings, it is evident that the indoor environment is significantly impacted by outdoor conditions. These conditions vary widely, spanning local weather phenomena like cold temperatures that cause pollutants to accumulate at lower altitudes, to distant events such as Sahara sands being carried thousands of kilometres by the wind to Norway.

The correlation between indoor and outdoor PM10 values varies significantly across locations and classrooms. This variation may indicate diverse sources of particulate matter, seasonal influences on indoor air quality, or differences in ventilation systems and cleaning procedures. To delve deeper, machine learning was employed using a random forest regression algorithm. This algorithm utilized various outdoor and indoor parameters to determine feature importance and predict indoor PM10 levels. The models demonstrated strength, with some achieving R^2 values as high as 0.92.

In January, outdoor temperature emerged as the most influential feature, closely followed by outdoor PM2.5 values and indoor relative humidity. Conversely, in September, indoor relative humidity held the most influence. Indoor CO_2 levels had influence in both January and September, likely attributed to consistent student activity.

Given the robustness of the random forest models and the close alignment between the predictive models and the actual PM10 values, there exists a high potential for future cleaning procedures to be founded upon predictive models. Embracing modern technology in cleaning processes could reduce redundant work and significantly enhance the indoor environment.

6 REFERENCES

Berglund, B. a. (1992). Effects of indoor air pollution on human health. *Indoor air*, 2--25.

Eichholtz, P. a. (2024). The effect of post-COVID-19 ventilation measures on indoor air quality in primary schools. *PNAS nexus*, pgad429.

Fenger, J. (2009). Air pollution in the last 50 years--From local to global. *Atmospheric environment*, 13--22.

Wei, W. a. (2019). Machine learning and statistical models for predicting indoor air quality. *Indoor Air*, 704--726.

World Health Organization. (2010). *WHO guidelines for indoor air quality: selected pollutants.* World Health Organization Regional Office for Europe.: World Health Organization.