# Towards Data-Driven Low-Carbon Concrete: Development and Implementation of the Concrete Mix Database (CDB)

Anıl Esen

DETOC Programme, C2CA Technology, RWTH Aachen University

Corresponding author: [aesen@c2ca.tech](mailto:aesen@c2ca.tech), [anil.esen@rwth-aachen.de](mailto:anil.esen@rwth-aachen.de)

ORCID: 0009-0005-3009-3388

**Abstract**

The urgent need to reduce the environmental impact of concrete while maintaining or improving its performance necessitates a data-driven approach to mix design and optimization. However, concrete mix design data in scientific literature is characterized by significant heterogeneity in terminology, calculation methods, testing procedures, and reporting formats. The water-to-binder ratio, a critical parameter for performance prediction, is inconsistently defined across studies—variously incorporating different supplementary cementitious materials (SCMs) and accounting differently for water from admixtures and aggregate absorption. Similarly, compressive strength results are reported using diverse specimen geometries and testing ages, complicating direct comparisons. This paper presents the Concrete Mix Database (CDB), a comprehensive relational database framework specifically designed to address these challenges through standardized terminology, normalized material properties, and structured relationships between concrete constituents and performance measures.

The CDB implements a component-based mix representation with explicit material hierarchies, allowing precise tracking of cement, SCMs, aggregates (both natural and recycled), admixtures, and fibers. A unified water-binder ratio calculation methodology has been established, with differentiated fields for water-to-cement ratio, water-to-binder ratio, and effective water-to-binder ratio with k-value considerations. Performance results are categorized by type (fresh, hardened, durability) and linked to specific testing methods, specimens, and curing conditions, facilitating meaningful comparison across datasets. The database structure also incorporates specialized extensions for recycled aggregates and sustainability metrics, supporting research on low-carbon concrete formulations.

Initial implementation has integrated data from six major literature sources (DS1-DS6), encompassing high-performance concrete, recycled aggregate concrete, self-consolidating concrete, and mixes with various SCMs. During this integration, dataset-specific challenges were systematically addressed, including variations in water content accounting, strength specimen normalization, and SCM reactivity quantification. Data quality assurance protocols have been implemented to flag assumptions, maintain traceability to source literature, and provide confidence scores for derived values.

The CDB enables analytical capabilities for concrete research, supporting material efficiency optimization, fine-grained analysis of SCM effects, recycled material performance prediction, and sustainability assessment. By providing a standardized, machine learning-ready data structure with clear relationships between composition and performance, the database lays the foundation for accelerated development of sustainable concrete technologies. The framework includes a contribution system designed to grow over time through community engagement, with rigorous validation protocols to maintain data integrity. WIP (Work In Progress)-.

**Keywords**: Concrete database; Materials informatics; Supplementary cementitious materials; Recycled concrete aggregates; Water-binder ratio; Low-carbon concrete; Machine learning

**Note**: This abstract represents work in progress (DC6) as part of the DETOCS Programme. The database structure and integration methodologies described are being actively refined as additional datasets are incorporated and analytical capabilities are expanded.

# 1. Introduction

## 1.1 Background and Motivation

* Current challenges in concrete sustainability and performance optimization
* Role of data-driven approaches in concrete mix design
* Need for comprehensive, standardized concrete mix databases

## 1.2 Literature Review: Existing Databases and Their Limitations

* Review of previous compilation efforts (including DS1-DS6 approaches)
* Identification of key limitations:
  + Inconsistent terminology and calculation methods
  + Varying specimen geometries and testing methods
  + Limited material characterization
  + Insufficient handling of SCMs and recycled materials

## 1.3 Research Objectives

* Development of a comprehensive database structure for concrete mix designs
* Standardization of key parameters (w/c and w/b ratios, strength values, etc.)
* Integration of diverse data sources
* Support for advanced analytics and machine learning
* Facilitation of sustainable concrete development

## 1.4 Paper Organization

* Brief outline of the subsequent sections

# 2. Database Design Methodology

## 2.1 Domain Analysis and Requirements

* Core concrete domain concepts and their relationships
* Key analytical workflows in concrete science
* Requirements gathering process
* Primary user stories and use cases

## 2.2 Database Architecture Overview

* Core entity structure and relationships
* Entity-relationship diagram
* Hierarchical schema design decisions

## 2.3 Key Design Decisions

* Material classification system and extension approach
* Component-based mix representation
* Performance result categorization
* Reference architecture (standards, methods, bibliographic entries)

## 2.4 Data Standardization Framework

* Terminology standardization (especially for water-binder ratios)
* Compressive strength normalization approach
* Material property standardization
* Units management and conversion

# 3. Database Schema Definition

## 3.1 Core Entities

* Concrete mix representation
* Material classification hierarchy
* Mix component system
* Performance results framework
* Sustainability metrics

## 3.2 Supporting Reference Tables

* Material properties dictionary
* Standards and test methods
* Units and conversions
* Bibliographic references

## 3.3 Material-Specific Extensions

* Cement-specific properties
* SCM characterization
* Aggregate characterization (natural and recycled)
* Admixture properties
* Fiber characterization

## 3.4 Specialized Performance Categories

* Fresh concrete properties
* Hardened concrete properties
* Durability properties
* Sustainability indicators

# 4. Data Integration Methodology

## 4.1 Source Dataset Selection Criteria

* Literature review methodology
* Inclusion/exclusion criteria
* Dataset quality assessment

## 4.2 Data Extraction and Transformation Process

* Automated and manual extraction methods
* Handling of data gaps and assumptions
* Reconciliation of terminological differences
* Conversion between different testing standards

## 4.3 Dataset-Specific Challenges and Solutions

* **DS1 (High-Performance Concrete)**
  + Handling of unreported fly ash classes
  + Superplasticizer details standardization
  + Standardization to 15-cm cylinders
* **DS2 (Recycled Concrete Aggregates)**
  + "Effective w/c ratio" interpretation
  + Handling of digitized data
  + RCA variability documentation
* **DS3 (Recycled Aggregate Concrete - ML)**
  + Dataset balancing issues
  + Conversion to standardized units
  + Specimen size standardization
* **DS4 (Self-Consolidating Concrete)**
  + "Water to powder ratio" conversion
  + Handling of limestone powder inclusion in binder
  + Missing test results management
* **DS5 (GGBS and Fly Ash Concretes)**
  + Cement strength class assignment for non-European literature
  + Normalization of varying curing conditions
  + Model bias correction for different SCMs
* **DS6 (Recycled Aggregate Concrete with SCMs)**
  + Implementation of "effective w/b ratio" calculation
  + Superplasticizer water content assumptions
  + Integration of oxide-based reactivity moduli

## 4.4 Data Quality Assurance and Validation

* Validation methodology
* Cross-checking procedures
* Anomaly detection
* Confidence scoring system

# 5. Current Database Status and Content

## 5.1 Overview of Integrated Datasets

* Summary statistics of current database state
* Distribution by material types
* Performance range coverage
* Geographical and temporal distribution

## 5.2 Water-Binder Ratio Standardization

* Unified definition implementation
* Comparative analysis of different calculation methods
* Impact on strength prediction
* Special considerations for SCMs and k-value concept

## 5.3 Compressive Strength Normalization

* Conversion factors between specimen types
* Statistical validation of conversion approach
* Age-dependent strength relationships
* Curing condition effects

## 5.4 Material Characterization Depth

* Chemical composition coverage
* Physical properties representation
* Special focus on SCMs and recycled materials
* Property-performance relationship mapping

# 6. Database Applications and Case Studies

## 6.1 Conventional Concrete Analysis

* Water-binder ratio vs. strength relationships
* Age-dependent strength development
* Mix optimization workflows
* Regional variation analysis

## 6.2 SCM-Containing Concrete Analysis

* SCM replacement level effects
* SCM type comparison
* Synergistic effects of multiple SCMs
* k-value concept validation

## 6.3 Recycled Aggregate Concrete Analysis

* RCA replacement level impact
* RCA quality influence
* Interaction between RCAs and SCMs
* Predictive models for RAC performance

## 6.4 Machine Learning Applications

* Feature importance analysis
* Model performance comparison on standardized data
* Transfer learning between concrete types
* Performance prediction accuracy improvements

## 6.5 Sustainability Optimization

* Carbon footprint vs. performance trade-off analysis
* Material efficiency optimization
* Regional availability consideration
* Service life incorporation

# 7. Future Development and Community Engagement

## 7.1 Database Evolution Roadmap

* Planned feature additions
* Schema extension capabilities
* Integration of additional datasets
* Linkage to other material databases

## 7.2 Data Contribution Framework

* Standardized submission process
* Quality control for contributed data
* Validation workflows
* Attribution and credit system

## 7.3 API and Tool Development

* Programmatic access capabilities
* Integration with common analysis tools
* Visualization frameworks
* Machine learning pipelines

## 7.4 Community Building and Governance

* Access and usage policies
* Contribution guidelines
* Expert review panels
* Sustainability of the initiative

# 8. Conclusions and Outlook

## 8.1 Summary of Contributions

* Major innovations in concrete data standardization
* Integration achievements
* Analytical capabilities enabled
* Research questions addressed

## 8.2 Limitations and Future Work

* Current data gaps
* Methodological limitations
* Potential extensions
* Future research directions

## 8.3 Broader Impact

* Potential influence on concrete sustainability
* Educational applications
* Industry adoption pathways
* Policy implications

**Acknowledgments**

* DETOCS Programme recognition
* Contributors and reviewers
* Funding sources

**References**

* Literature sources (including DS1-DS6 papers)
* Methodological references
* Related database initiatives
* Technical standards

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.